



**Application des algorithmes d'apprentissage automatique pour la détection de défauts de roulements sur les machines tournantes dans le cadre de l'Industrie 4.0**

**par WendBenedo Arnaud Bienvenue ZOUNGRANA**

**Mémoire présenté à l'Université du Québec à Chicoutimi en vue de l'obtention du grade de Maître ès sciences appliquée en ingénierie, profil recherche**

Québec, Canada

© WendBenedo Arnaud Bienvenue ZOUNGRANA, 2020

## RÉSUMÉ

L'internet des objets industriels (IIoT) fait partie d'un concept plus large connu sous le nom de L'Internet des Objets, ou IdO (en anglais Internet of Things, ou IoT). Les IIoT apportent de nouvelles opportunités aux sites de production telles que la diminution des coûts des opérations et l'augmentation de la productivité dans le but d'une exploitation optimale. La technologie IIoT révolutionnera les procédés de fabrication industrielle en permettant l'acquisition des quantités importantes de données, à des vitesses beaucoup plus élevées, et bien plus efficaces qu'auparavant. Un certain nombre d'entreprises innovantes ont commencé à mettre en œuvre l'IIoT en exploitant des appareils connectés intelligents dans leurs usines (c'est ce qu'on appelle les usines intelligentes ou Industrie 4.0). Dans une perspective d'acquisition des données, l'internet des objets a favorisé l'inclusion des sous-systèmes ainsi que leurs analyses en temps réel. Pour ce faire, l'Industrie 4.0 introduit un concept de production numérisée en permettant une intégration souple et agile de nouveaux modèles commerciaux tout en maintenant les coûts de fabrications et l'efficacité à un niveau raisonnable.

Dans ce projet de recherche, nous allons étudier la maintenance prédictive des installations industrielles. Cette tâche est essentielle au bon fonctionnement de l'usine et à la sécurité des personnes. Compte tenu des coûts, il est judicieux d'établir un bon équilibre entre entretien préventif systématique et entretien correctif. La surveillance des installations concourt à limiter le niveau d'entretien préventif. Dans ce contexte, l'analyse vibratoire constitue un outil de détection puis de diagnostic de défauts de fonctionnement des installations. Aussi, après avoir décrit les principales manifestations vibratoires des défauts de fonctionnement des machines, nous allons examiner les stratégies de détection et de surveillance dans le domaine temporel et fréquentiel et la démarche de diagnostic en nous appuyant sur l'intelligence artificielle et en analysant les signaux vibratoires permettant de déduire une politique de gestion de la maintenance.

Notre objectif principal est de réaliser un système permettant d'assurer l'analyser les signaux vibratoires d'une machine tournante dans le domaine temps/fréquence. Il sera ensuite aisé de le comparer avec un système d'apprentissage automatique capable de détecter et classer les défauts grâce à des algorithmes d'intelligence artificielle. L'application vise à fournir un système de détection de défauts fiable afin de réduire les temps de dépannages et favoriser un diagnostic rapide des pannes des systèmes industriels. Par conséquent, le projet va de la conception jusqu'à la mise en œuvre des algorithmes informatiques avec des exemples réels de signaux vibratoires. Un processus d'optimisation sera mis en œuvre lors de la prise des décisions de l'équipe humaine afin d'augmenter l'efficacité des résultats et réduire les situations à risque qui mettront les machines hors d'usage. Ce projet de recherche permettra donc d'introduire un système intelligent dans un environnement de production de l'industrie 4.0.

Durant cette étude, nous avons implémenté dans un premier temps des algorithmes qui nous ont permis d'extraire des caractéristiques des signaux d'une machine tournante. Ensuite, nous avons mis en place un système de surveillance de l'état de cette machine en fixant un seuil pour le bon fonctionnement et un autre pour déclencher une alarme quand ce dernier est atteint. Dans un second temps, nous avons utilisé des algorithmes d'apprentissage automatique (ou machine learning) pour classer les différents niveaux de défaillance. Après extraction des caractéristiques des signaux dans le domaine temporel et fréquentiel nous avons obtenu une fiabilité de 99.3% avec la méthode d'estimation dite « validation croisée » (en anglais cross-validation). Ce processus d'apprentissage optimise les paramètres du modèle afin que celui-ci corresponde aux données le mieux possible. Ensuite, nous avons évalué une autre technique de validation « testset validation » (en anglais holdout method). Cette technique est recommandée pour les grands ensembles de données. Après plusieurs tests, nous avons pu obtenir un taux de classification de 100% pour les différents niveaux de défauts considérés.

## ABSTRACT

The internet of industrial objects (IIoT) is part of a larger concept known as the "Internet of Things (IoT)". IoT's bring new opportunities to production sites such as lower operating costs and increased productivity for optimal operation. The application of IoT to the manufacturing industry is called IIoT (or Industrial Internet or Industry 4.0). The IoT will revolutionize manufacturing by enabling the acquisition and accessibility of massive amounts of data at much higher speeds and far more efficiently than before. A few innovative companies have started to implement IoT by exploiting smart connected devices in their factories (so-called smart factories or Industry 4.0). From a data acquisition perspective, the Internet of Things has favoured the inclusion of subsystems and their analysis in real time. To achieve this, Industry 4.0 introduces a digitized production concept by allowing flexible and agile integration of new business models while keeping manufacturing costs and efficiency at a reasonable level.

In this thesis, we will study the predictive maintenance for industry 4.0. This method of preventing asset failure by analyzing production data to identify patterns and predict issues before they happen. However, considering the costs, it is wise to strike a better balance between routine preventative maintenance and corrective maintenance. Facility monitoring helps to reduce the level of preventative maintenance. In this context, vibration analysis is a tool for detecting and then diagnosing malfunctions of installations. Also, after describing the main vibratory manifestations of failures of the machines, we will examine the detection and surveillance strategies at the time and frequency domain and the diagnosis process based on artificial intelligence by analyzing the vibratory signals.

Our main objective is to design a system that will analyze the vibrating signals of a rotating machine in the time/frequency domain. A machine learning system will be used to detect and classify the defects based on artificial intelligence algorithms. The application aims to provide a reliable fault detection system to reduce repair times and promote rapid diagnosis of industrial operations. Therefore, the project goes from the design to the implementation of new algorithms using vibratory signals. An optimization process will be implemented when the decisions of the adequate staff are made to increase the efficiency of the results and reduce the risk situations that will put the machines out of use. Therefore, this research project will introduce an intelligent system into an industry 4.0 production environment.

During this research project, firstly, we will implement algorithms that allowed us to extract characteristics of the vibration signals and set up a system to monitor the state of a rotating machine by setting a threshold of operation and trigger an alarm when this threshold is reached. In a second step, we will use signal processing and machine learning toolkits to classify the different levels of machine failure and use this method to detect the presence of error when the machine is running. After extracting the characteristics of our signals at the time and frequency domain, we will be obtained with the "cross-validation" a recognition rate of 99.3%. The final implementation with the "holdout validation" recommended for large datasets allowed us to have a classification rate of 100% of the different levels of defects considered.

## TABLE DES MATIÈRES

RÉSUMÉ .....	i
ABSTRACT .....	ii
TABLE DES MATIÈRES .....	iii
LISTE DES TABLEAUX .....	vi
LISTE DES FIGURES .....	vii
LISTE DES SIGLES ET ACRONYMES .....	viii
REMERCIEMENTS .....	ix
<b>CHAPITRE 1 .....</b>	<b>1</b>
<b>INTRODUCTION .....</b>	<b>1</b>
1.1. CONTEXTE .....	1
1.2. PROBLÉMATIQUE .....	2
1.3. MÉTHODOLOGIE .....	3
1.4. ORGANISATION DU MÉMOIRE .....	4
<b>CHAPITRE 2 .....</b>	<b>5</b>
<b>CADRE THÉORIQUE ET REVUE DE LA LITTÉRATURE .....</b>	<b>5</b>
2.1. INTERNET DES OBJETS (IoT) .....	5
2.1.1. DÉFINITION .....	5
2.1.2. CONCEPT DE L'IoT .....	6
2.1.3. ARCHITECTURE ET PROTOCOLE DES IoT .....	7
2.1.4. SÉCURITÉ DES RÉSEAUX IoT .....	11
2.1.5. DOMAINE D'APPLICATION DES IoT .....	12
2.2. CLOUD COMPUTING .....	13
2.3. BIG DATA .....	14
2.4. INDUSTRIE 4.0 .....	14
2.5. TYPES DE MAINTENANCE .....	16
2.5.1. LA MAINTENANCE CORRECTIVE .....	16
2.5.2. LA MAINTENANCE PRÉVENTIVE .....	17
2.5.3. LA MAINTENANCE PRÉDICTIVE .....	17
2.6. ANALYSE ET CLASSIFICATION DE DONNÉES .....	17
2.6.1. ANALYSE DE DONNÉE .....	18
2.6.2. CLASSIFICATION DE DONNÉES .....	22
2.6.2.1. APPRENTISSAGE SUPERVISÉ .....	24
2.6.2.2. APPRENTISSAGE NON SUPERVISÉ .....	25
2.7. REVUE DE LA LITTÉRATURE .....	25

<b>CHAPITRE 3.....</b>	<b>27</b>
<b>MÉTHODOLOGIE.....</b>	<b>27</b>
3.1. INTRODUCTION.....	27
3.2. PROTOCOLE D'ACQUISITION DES SIGNAUX.....	27
3.3. PRÉTRAITEMENT DES SIGNAUX.....	29
3.4. SÉLECTION DES CARACTÉRISTIQUES.....	31
3.4.1. CARACTÉRISTIQUES TEMPORELLES.....	31
3.4.2. CARACTÉRISTIQUES FRÉQUENTIELLES.....	34
3.4.3. CARACTÉRISTIQUES TEMPS-FRÉQUENCE.....	37
3.4.4. LES LIMITES DES APPROCHES TEMPS-FRÉQUENCE.....	37
3.5. SYSTÈME DE SURVEILLANCE D'ÉTAT DE LA MACHINE.....	37
3.6. LES ALGORITHMES DE CLASSIFICATION.....	38
3.6.1. ARBRES DE DÉCISION.....	39
3.6.2. MACHINES A VECTEURS DE SUPPORT (SVM).....	41
3.6.3. LES k-PLUS PROCHES VOISINS (k-NN).....	43
3.6.4. NAÏVE BAYES.....	45
3.6.5. RÉSEAUX NEURONES ARTIFICIELS (ANN).....	47
3.7 CONCLUSION.....	50
<b>CHAPITRE 4.....</b>	<b>51</b>
<b>ANALYSE DES RÉSULTATS ET DISCUSSION.....</b>	<b>51</b>
4.1. APERÇU DES SIGNAUX.....	51
4.2. ANALYSE DES CARACTÉRISTIQUES SÉLECTIONNÉES.....	52
4.2.1. MÉTHODE TEMPS-FREQUENCE.....	52
4.2.2. MÉTHODE STATISTIQUE.....	54
4.3. RÉSULTATS DES TESTS.....	55
4.3.1. RÉSULTATS DU SYSTÈME DE SURVEILLANCE DE L'ÉTAT DE LA MACHINE.....	56
4.3.2. RÉSULTATS DES TESTS DE CLASSIFICATION DES SIGNAUX.....	57
4.3.2.1. MÉTHODE DE VALIDATION CROISÉE (K-FOLD).....	58
4.3.2.2. MÉTHODE DE VALIDATION HOLDOUT.....	60
4.3.2.3. MATRICE DE CONFUSION.....	61
<b>CHAPITRE 5.....</b>	<b>65</b>
<b>CONCLUSION ET PERSPECTIVES.....</b>	<b>65</b>
5.1. CONCLUSION.....	65
5.2. PERSPECTIVE.....	66
BIBLIOGRAPHIE OU LISTE DE RÉFÉRENCES.....	67
ANNEXES.....	ix

ANNEXE 1 : ARTICLE SCIENTIFIQUE .....	ix
<b>1 Introduction .....</b>	<b>x</b>
<b>2 Methodology .....</b>	<b>xi</b>
2.1 Data Acquisition Protocol .....	xi
2.2 Data pre-processing .....	xii
2.3 Time signal based approach .....	xii
<b>3 Predictive maintenance of rotating machines .....</b>	<b>xiv</b>
3.1 Classical machine learning based approaches .....	xv
3.2 Support Vector Machines (SVM).....	xv
3.3 Decision Trees .....	xv
3.4 k-Nearest Neighbour (k-NN) Classifier.....	xv
3.5 Naive Bayes classifier .....	xv
<b>4 Results and Discussion .....</b>	<b>xv</b>
4.1 Test results with cross-validation .....	xvi
<b>5 Conclusion .....</b>	<b>xvii</b>
<b>References .....</b>	<b>xviii</b>
ANNEXE 2 : PLATEFORME DE DÉVELOPPEMENT MATLAB .....	xix
ANNEXE 3 : CALCULS STATISTIQUES.....	xx

## LISTE DES TABLEAUX

Tableau 1 : Comparaison des types d'analyses de classification.....	23
Tableau 2 : Caractéristiques temporelles .....	32
Tableau 3 : Caractéristiques fréquentielles.....	35
Tableau 4 : Résultats des tests avec la validation croisée .....	64
Tableau 5 : Résultats des tests avec la validation holdout.....	64

## LISTE DES FIGURES

Figure 1 : Architecture des réseaux IoT.....	8
Figure 2 : Étapes d'analyse de donnée (classification) [29]. ....	19
Figure 3 : Taxonomie des modèles d'apprentissage automatique [32]. ....	23
Figure 4 : Banc d'essai et illustrateur de placement des capteurs [9]. ....	28
Figure 5 : Signal de vibration de la machine enregistré jusqu'à la défaillance.....	29
Figure 6 : Découpage du signal en trois groupes selon la présence ou non de défauts. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant. ....	30
Figure 7 : Spectre de kurtosis dans le domaine temporel des signaux. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.....	33
Figure 8 : Spectre d'amplitude des signaux obtenus de la FFT. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.....	36
Figure 9 : Exemple d'illustration d'un arbre de décision.....	40
Figure 10 : Principe de l'algorithme SVM.....	42
Figure 11 : Organigramme de l'algorithme SVM.....	42
Figure 12 : Principe de l'algorithme k-NN. ....	44
Figure 13 : Organigramme de l'algorithme KNN. ....	44
Figure 14 : Principe de l'algorithme Naïve Bayes. ....	46
Figure 15 : Organigramme de l'algorithme Naïve Bayes. ....	46
Figure 16 : Principe des réseaux de neurones artificiels.....	48
Figure 17 : Organigramme de l'algorithme ANN. ....	49
Figure 18 : Fréquences maximales moyennes extraites en fonction du temps. ....	51
Figure 19 : Tracer du spectrogramme des signaux de vibration. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.....	53
Figure 20 : Histogramme des caractéristiques des signaux de vibration.....	54
Figure 21 : Mise en œuvre du système de surveillance avec détection de panne. ....	56
Figure 22 : Classification des signaux avec la méthode de validation croisée. ....	59
Figure 23 : Classification des signaux avec la méthode de la validation holdout. ....	61
Figure 24 : Matrice de confusion avec la méthode de validation croisée. ....	62
Figure 25 : Matrice de confusion avec la méthode de la validation holdout. ....	63



## LISTE DES SIGLES ET ACRONYMES

<b>ACP</b>	Analyse en Composantes Principales
<b>ANN</b>	Artificial Neural Network
<b>BPFI</b>	Ball Pass Frequency Inner
<b>BPFO</b>	Ball Pass Frequency Outer
<b>BSF</b>	Ball Spin Frequency
<b>FFT</b>	Fast Fourier Transform
<b>GMAO</b>	Gestion de Maintenance Assistée par Ordinateur
<b>GPU</b>	Graphics Processing Unit
<b>HaaS</b>	Hardware as a Service
<b>IoT</b>	Internet of Things
<b>IIoT</b>	Industrial Internet of Things
<b>IdO</b>	Internet des objets
<b>IMS</b>	Intelligent Maintenance Systems
<b>IUCRC</b>	Industry-University Cooperative Research Centers
<b>KNN</b>	K-Nearest Neighbors
<b>MATLAB</b>	Matrix Laboratory
<b>NASA</b>	National Aeronautics and Space Administration
<b>NoSQL</b>	Not only Structured Query Language
<b>NSF</b>	National Science Foundation
<b>RUL</b>	Remaining Useful Life
<b>RMS</b>	Root Mean Square
<b>SaaS</b>	Software as a Service
<b>SVM</b>	Support Vector Machine
<b>SVR</b>	Support Vector Regression
<b>UIT</b>	Union Internationale des Télécommunications
<b>UQAC</b>	Université du Québec à Chicoutimi

## REMERCIEMENTS

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Je voudrais tout d'abord adresser toute ma reconnaissance à mon directeur de mémoire, professeur Abdellah Chehri, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion.

Je désire aussi remercier les professeurs de l'Université du Québec à Chicoutimi (UQAC), qui m'ont fourni les outils nécessaires à la réussite de mes études universitaires, mes pensées vont à l'endroit de professeure Suzy Ngomo.

Je tiens à remercier spécialement professeur Martin Otis, qui fut mon premier directeur de recherche et qui m'a offert la formidable opportunité de rejoindre son équipe.

Un grand merci à professeure Salmata Ouedraogo qui m'a permis de découvrir l'UQAC grâce à son formidable travailleur de recrutement, ses conseils et son soutien permanent tout au long de mon travail.

Je voudrais exprimer ma reconnaissance à toute ma famille qui me pousse à aller de l'avant et qui a toujours été là pour moi. Je vous suis reconnaissant et je vous dédie ce mémoire. Merci pour votre confiance et votre soutien inestimable.

Enfin, je tiens à témoigner toute ma gratitude envers les amis et collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de ma démarche.

## **CHAPITRE 1**

### **INTRODUCTION**

#### **1.1. CONTEXTE**

Devant le rythme effréné de l'adoption des technologies numériques, les organisations cherchent à se transformer rapidement. Les nouvelles technologies, les nouveaux produits et services, de même que les nouveaux modèles d'affaires peuvent perturber les opérations. Dans ce cas, l'adoption des principes de l'Industrie 4.0 devient une nécessité. Les dirigeants de demain doivent être prêts à se doter d'une nouvelle structure organisationnelle [1-5].

L'Industrie 4.0 est une initiative du gouvernement allemand visant à transformer le futur secteur manufacturier en Europe en renforçant la numérisation et l'interconnexion entre produits, processus et modèles économiques. La transformation des systèmes automatisés et robotiques (Industrie 3.0) en systèmes cyberphysiques restructure numériquement l'ensemble du processus de fabrication et de la chaîne de valeur [6].

La transformation des actifs automatisés en actifs cyberphysiques est une tâche complexe. Cette transformation numérique des actifs et des opérations intelligentes sera impossible sans la mise en œuvre d'un système intelligent de gestion informatisée de la maintenance (GMAO), un élément essentiel du système de gestion de la maintenance. Cela exige de passer de la maintenance préventive à la maintenance prédictive, où une panne de machine pourrait être prévenue avant qu'elle ne se produise. Cette transformation pourrait être réalisée grâce aux données collectées par des capteurs intelligents qui pourraient être analysées et utilisées pour la planification de la maintenance. Cependant, il faudra développer des capacités informatiques et cognitives dans la GMAO pour utiliser ces données en temps réel pour la prise de décision en matière de maintenance [7].

## 1.2. PROBLÉMATIQUE

La maintenance traditionnelle s'opère habituellement après le dysfonctionnement d'un équipement ou d'un appareil. Cette approche réactive consistant à décrire les dysfonctionnements d'un équipement ou d'un appareil une fois qu'ils se sont produits et à les traiter a posteriori, cela représente le pire des scénarios en matière de maintenance. Dans le cadre de la maintenance préventive, les opérateurs ont la possibilité d'exécuter des tâches de maintenance en continu.

Le modèle de rupture qui est la maintenance prédictive exploite pleinement la puissance de l'Internet des objets, les mégadonnées (en anglais Big Data) et l'infonuagique (en anglais Cloud Computing) de façon à modifier la dynamique de la maintenance productive totalement conventionnelle. Ce modèle définit la maintenance comme le simple fait de réduire les périodes d'immobilisation des machines. L'intégration de logiciels intelligents aux machines de production facilite les diagnostics et signale les problèmes techniques à l'avance. L'intégration de logiciels intelligents renforce la proactivité dans la maintenance.

Les analyses prédictives identifient de potentielles options à privilégier à des fins d'anticipation. La maintenance prédictive s'appuie quant à elle sur l'ensemble de ces approches et de leurs fonctions. Un tel type de maintenance englobe à la fois les événements susceptibles de survenir et la mise en œuvre de stratégies de maintenance optimisées. Avec la maintenance prédictive, les appareils participent de façon proactive à leur propre maintenance en collaboration avec les opérateurs [8].

### 1.3. MÉTHODOLOGIE

La méthodologie proposée pour la réalisation de ce projet est décrite en trois étapes. La première étape consiste à recueillir des données par les capteurs placés sur un équipement en production durant son cycle de vie, c'est-à-dire du bon fonctionnement de la machine jusqu'à ce que surviennent les cas de défaillance. Pour les besoins de ce projet nous avons utilisé des données fournies par le site de la National Aeronautics and Space Administration (NASA) dans le référentiel de dépôt de données de pronostics au cours d'une expérience sur les roulements d'une machine tournante. L'ensemble des données a été fourni par le Center for Intelligent Maintenance Systems (IMS) de l'Université de Cincinnati [9].

Dans la deuxième étape, nous avons procédé à l'analyse et au prétraitement des signaux afin d'en sélectionner les meilleurs qui permettront de réaliser nos objectifs c'est-à-dire la détection de panne et de distinguer une meilleure classification des signaux de bon fonctionnement et de défaut. Concernant la deuxième étape, nous avons utilisé le logiciel MATLAB pour la caractérisation des signaux. Nous avons également utilisé la combinaison des techniques de traitement temporel et statistique pour l'extraction des informations utiles.

Enfin, la dernière étape va consister dans un premier temps à fixer un seuil de bon fonctionnement de la machine et grâce à l'extraction de la moyenne des fréquences crêtes, nous allons suivre le fonctionnement de la machine dans l'intervalle des seuils et nous assurer de son bon fonctionnement. Ensuite dans un second temps, nous allons utiliser les méthodes de classifications basées sur l'apprentissage automatique pour la reconnaissance des différents niveaux de fonctionnement de la machine. Le but est de reconnaître les bons et les mauvais signaux à partir des informations extraites des étapes précédentes.

#### **1.4. ORGANISATION DU MÉMOIRE**

Ce mémoire a été organisé selon les directives fournies par l'Université du Québec à Chicoutimi. Le mémoire a été divisé en cinq chapitres en plus de la conclusion.

Le premier chapitre, qui fait figure d'introduction générale, son but est de mettre en évidence le contexte et la problématique du sujet. Dans le chapitre 2, nous avons présenté les théories relatives à l'IoT, l'infonuagique et les données massives dans l'Industrie 4.0. Au chapitre 3, nous avons décrit les différentes méthodes d'analyse et de classification de données. Et enfin, au chapitre 4 nous avons présenté la discussion des résultats et pour terminer au chapitre 5 avec la conclusion et les perspectives de notre projet de recherche.

## **CHAPITRE 2**

### **CADRE THÉORIQUE ET REVUE DE LA LITTÉRATURE**

#### **2.1. INTERNET DES OBJETS (IoT)**

##### **2.1.1. DÉFINITION**

L'Internet des objets (IoT) permet à des objets et des appareils qui jadis fonctionnaient de manière isolée et non connectée de se détecter, de communiquer, d'interagir et de collaborer au sein d'un même réseau.

Selon la définition de l'Union internationale des télécommunications (UIT) « L'Internet des objets : est une infrastructure mondiale pour la société de l'information, qui permet de disposer de services évolués en interconnectant des objets (physiques ou virtuels) grâce aux technologies de l'information et de la communication interopérables existantes ou en évolution » [10].

L'Internet des objets se veut l'avenir de l'Internet actuel. Il est généralement défini comme un réseau de périphériques ou d'objets physiques et virtuels capables de collecter des données environnantes et de les échanger entre eux ou via Internet. Pour permettre la collecte de données, il faut intégrer aux périphériques des capteurs, des logiciels et des composants électroniques ; les échanges sont rendus possibles en les connectant à des réseaux locaux ou à Internet. Les origines de l'Internet des objets sont diffuses. Bien que le mot ait été inventé pour la première fois en 1999 par Kevin Ashton, cofondateur et directeur exécutif du centre Auto-ID du MIT, pour des entreprises telles que CISCO, l'IoT est né en 2009, lorsque plus d'appareils que de personnes étaient connectés au réseau internet. À cette époque, le nombre

d'appareils connectés s'élevait à 10 milliards, mais les attentes sont nombreuses. On pense que d'ici 2020, plus de 50 milliards d'appareils seront connectés à Internet [11].

La prolifération des capteurs et des puces à bas prix permet de concevoir des systèmes électroniques moins coûteux et permet la communication des données entre les objets connectés. En plus de cela, nous avons les différentes évolutions rapides dans le domaine des réseaux sans fil et de l'infrastructure infonuagique (Cloud) qui propose des solutions et des plateformes avancées spécialement conçu pour l'internet des objets. L'ensemble de ces facteurs influencent le développement de cette nouvelle étape de l'internet et la culture d'objet connecté au sein du grand public.

### **2.1.2. CONCEPT DE L'loT**

L'loT donne aux TIC une nouvelle dimension en ce sens que, grâce à lui, la communication est non seulement possible en tout lieu et à tout moment, comme c'était déjà le cas, mais aussi avec n'importe quel OBJET. Dans l'loT, les objets sont représentés par des objets du monde physique (objets physiques) ou du monde de l'information (objets virtuels), pouvant être identifiés et intégrés dans des réseaux de communication. Au fil du temps, le terme loT a évolué et est décrit comme un réseau d'entités connectées via n'importe quel type de capteur, permettant à ces entités, appelées objets physiques ou virtuels connectés à Internet, d'être localisées, identifiées et même exploitées.

Les objets physiques appartiennent au monde physique et peuvent être détectés, commandés et connectés. L'environnement qui nous entoure, les robots industriels, les biens et les équipements électriques sont autant d'exemples d'objets physiques. Et les objets virtuels appartiennent au monde de l'information ; on peut les stocker, les traiter et y accéder. Ces objets sont par exemple des contenus multimédias ou des logiciels [10].



### 2.1.3. ARCHITECTURE ET PROTOCOLE DES IoT

#### ➤ ARCHITECTURE

Dans cette partie, nous allons présenter l'architecture IoT et les protocoles généralement utilisés pour cette technologie. L'architecture IoT peut être considérée comme un système physique, virtuel ou qui regroupe les deux donc hybride.

Elle consiste ainsi à un regroupement d'ensembles d'objets physiques, de capteurs, de couches de communication, d'utilisateurs, et une couche application. Des architectures particulières agissent en tant que composant pivot d'une infrastructure spécifique à l'IoT, et les principales parties qui constituent cette architecture peuvent être regroupées en trois grands groupes. Nous avons d'abord la partie composée d'un parc d'objets connectés fixes ou mobiles, répartis géographiquement. Ensuite, nous avons la partie réseau qui va permettre de connecter les objets à internet ou entre eux pour transmettre les données, ce réseau peut être filaire ou sans-fil, court ou longue portée. Et enfin la partie application, développée le plus souvent en technologie web ou une application sur un téléphone intelligent qui collecte les données du réseau d'objets pour fournir une information traitée, compréhensible et qui permet de prendre une décision [12].

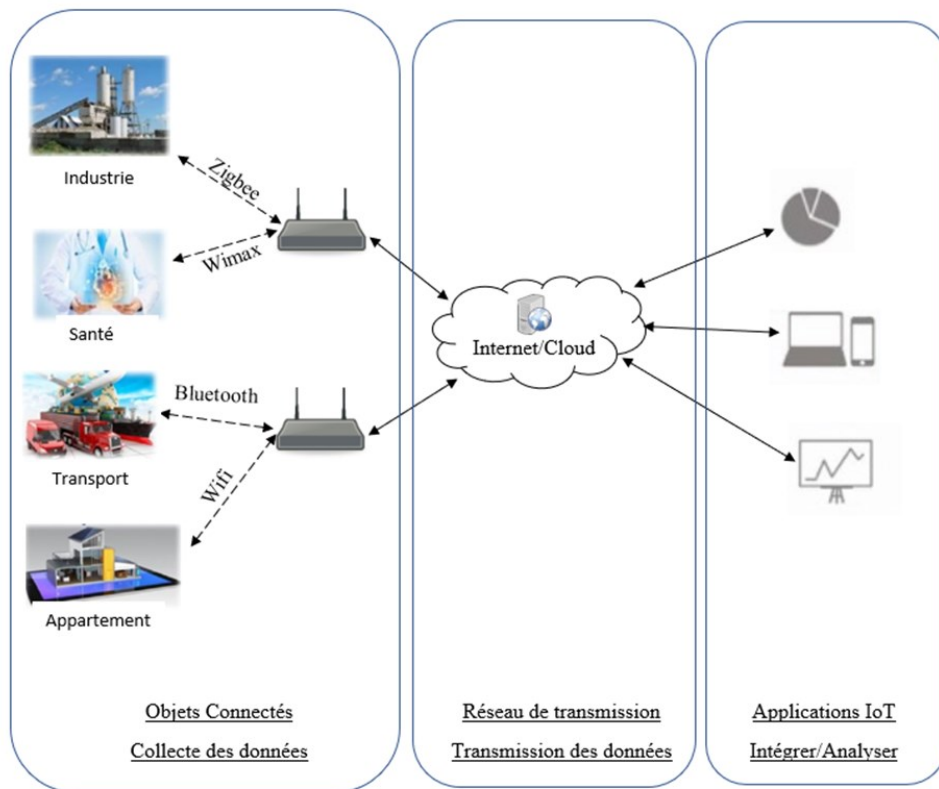


Figure 1 : Architecture des réseaux IoT

## ➤ PROTOCOLES

Les protocoles de communication constituent l'épine dorsale des systèmes IoT et permettent la connectivité au réseau et le couplage aux applications. Les protocoles de communication permettent aux appareils d'échanger des données via le réseau. Les protocoles définissent les formats d'échange de données, le codage des données, les schémas d'adressage pour les périphériques et le routage des paquets de la source à la destination. Les autres fonctions des protocoles incluent le contrôle de séquence, le contrôle de flux et la retransmission de paquets perdus. Dans cette partie, nous allons comparer différentes technologies de communication sans fil par rapport à divers paramètres [12].

✓ **802.11 – WiFi**

Le WiFi est basé sur la norme IEEE (802.11), c'est une collection de normes de communication pour les réseaux locaux sans fil (WLAN). Par exemple, le (802.11a) fonctionne dans la bande des 5 GHz, le (802.11b) et le (802.11g) dans la bande des 2,4 GHz, le (802.11n) dans les bandes des 2,4/5 GHz, le (802.11 ac) dans la bande des 5 GHz et le (802.11 ad) dans les bandes de 60GHz. Ces normes fournissent des débits allant de 1 Mb/s à 6,75 Gb/s. Le WiFi offre une portée de communication de l'ordre de 20 m (intérieur) à 100 m (extérieur).

✓ **802.16 - WiMax**

Le WiMax est basé sur la norme IEEE (802.16), c'est un ensemble de normes large bande sans fil. Les normes WiMAX fournissent des débits de données compris entre 1,5 Mb/s et 1 Gb/s. La mise à jour récente (802,16 m) fournit un débit de données de 100 Mb/s pour les stations mobiles et de 1 Gb/s pour les stations fixes. Les spécifications sont facilement disponibles sur le site Web du groupe de travail IEEE 802.16 [13].

✓ **802.15.4 - LR-WPAN**

LR-WPAN est basé sur la norme IEEE (802.15.4), c'est une collection de normes LR-WPAN (Wireless Personal Area Networks). Ces normes constituent la base des spécifications pour les protocoles de communication de haut niveau tels que ZigBee. Les normes LR-WPAN fournissent des débits allant de 40 kb/s à 250 kb/s. Ces normes permettent une communication à faible coût et à faible vitesse avec des périphériques soumis à des contraintes d'alimentation. Il fonctionne respectivement aux fréquences de 868/915 MHz et 2,4 GHz à des débits de données faibles et élevés. Les spécifications des normes (802.15.4) sont disponibles sur le site Web du groupe de travail IEEE802.15 [14].

### ✓ **2G / 3G / 4G - communication mobile**

Il existe différentes générations de normes de communication mobile, notamment les normes de deuxième génération (2G, notamment GSM et CDMA), de troisième génération (UMTS et CDMA2000 de troisième génération) et de quatrième génération (LTE, notamment de 4G). Les appareils IoT basés sur ces normes peuvent communiquer sur des réseaux cellulaires. Les débits de données pour ces normes vont de 9,6 kb/s (2G) à 100Mb/s (4G) et sont disponibles sur les sites Web de 3GPP.

### ✓ **802.15.1 - Bluetooth**

Bluetooth est basé sur la norme IEEE (802.15.1). Il s'agit d'une technologie de communication sans fil à faible consommation d'énergie qui convient à la transmission de données entre appareils mobiles sur une courte distance (8 à 10 m). La norme Bluetooth définit une communication PAN (Personal Area Network). Il fonctionne dans la bande 2,4 GHz. Le débit de données dans différentes versions du Bluetooth varie de 1 Mb/s à 24Mb/s. La version ultra basse consommation et à faible coût de cette norme porte le nom Bluetooth Low Energy (BLE ou Bluetooth Smart). Auparavant, en 2010, BLE avait fusionné avec la norme Bluetooth v4.0.

### ✓ **LoRaWAN R1.0 - LoRa**

LoRaWAN est un protocole de communication à longue portée récemment développé par LoRa<sup>™</sup> Alliance, une association ouverte et à but non lucratif. Son objectif principal est de garantir l'interopérabilité entre différents opérateurs dans une norme mondiale et ouverte. Les débits de données LoRaWAN vont de 0,3 kb/s à 50 kb/s. LoRa fonctionne dans les bandes ISM 868 et 900MHz. Selon Postscapes, LoRa communique entre les nœuds connectés dans un rayon de 20 miles (30 km), dans des environnements non obstrués. La durée de vie de la batterie dépend de l'utilisation du nœud.

#### **2.1.4. SÉCURITÉ DES RÉSEAUX IoT**

La tâche principale de l'IoT est d'interconnecter une multitude d'objets hétérogènes via Internet. Ces objets produiront un trafic beaucoup plus important sur le réseau, car davantage de données seront stockées. Par conséquent, une sécurité efficace pour l'IoT doit traiter de nombreux aspects tels que la confidentialité et la sécurité.

Certaines applications IoT prennent en charge des infrastructures sensibles et des services stratégiques tels que le réseau intelligent et la protection des installations. D'autres applications IoT généreront de plus en plus d'énormes quantités de données personnelles sur la situation des ménages, de la santé et des finances que les entreprises pourront exploiter pour leurs activités. L'absence de sécurité et de confidentialité créera une réticence à l'adoption de l'IoT par les entreprises et les particuliers. Il est possible de résoudre les problèmes de sécurité en formant les concepteurs de logiciels à intégrer des solutions de sécurité (systèmes de prévention des intrusions, pare-feu, par exemple) dans leurs produits et en encourageant les utilisateurs à utiliser les fonctionnalités de sécurité IoT intégrées à leurs appareils [15].

De plus en plus d'appareils connectés sont introduits dans les réseaux IoT, la menace potentielle pour la sécurité s'intensifie. Bien que l'IoT améliore la productivité des entreprises et améliore la qualité de vie des personnes, l'IoT augmentera également les surfaces d'attaque potentielles pour les pirates informatiques et autres cybercriminels. Une étude récente de Hewlett Packard (2014) a révélé que 70% des périphériques IoT les plus couramment utilisés contiennent de graves vulnérabilités. Les périphériques IoT présentent des vulnérabilités en raison d'un manque de cryptage de transport des données, d'interfaces Web non sécurisées, d'une protection insuffisante des logiciels et d'un niveau d'autorisation insuffisant. Les périphériques sur l'IoT n'utilisent généralement pas de techniques de cryptage de données[16].

### **2.1.5. DOMAINE D'APPLICATION DES IoT**

La disponibilité d'une large gamme de capteurs et d'appareils de différentes tailles et de capacités de calcul diverses fait de l'IoT un paradigme de premier plan pour permettre l'innovation dans différents domaines d'application. Les principaux domaines d'application des solutions IoT peuvent être regroupés comme suit : bâtiment [17, 18], environnement [19, 20], industrie, santé [21], transport et sécurité [22].

#### **➤ DOMAINE INDUSTRIEL**

Typiquement, le terme IoT industriel désigne l'utilisation de la technologie IoT pour la gestion du processus de fabrication. Dans ce contexte, l'IoT exploite les technologies de communication et d'automatisation de machine à machine (M2M) mise en œuvre depuis des années dans l'industrie. Les données acquises par les capteurs, par exemple les étiquettes RFID placées sur les articles, permettent aux entreprises de résoudre de manière rapide et efficace tous les problèmes rencontrés au cours du processus de la chaîne d'approvisionnement et de livraison, ce qui permet d'économiser du temps et de l'argent. Ce processus nécessite des conditions de travail restreintes, telles que des environnements soumis à des processus de travail à température contrôlée et à humidité contrôlée. Une autre application de la solution IoT dans le domaine industriel concerne le marché de la vente au détail. Par exemple, les technologies d'identification peuvent aider à lutter contre la contrefaçon en suivant tout le cycle de vie d'un produit, y compris des informations sur son emplacement et en précisant le processus de transport.

## ➤ **DOMAINE DE LA SÉCURITÉ**

L'utilisation de technologies compatibles avec l'IoT constitue une excellente application pour sécuriser les infrastructures critiques, telles qu'une infrastructure de production d'énergie, un stade, un port, une banque ou un autre environnement public. La surveillance par des capteurs peut fournir des informations sur la présence de personnes et leur comportement ou détecter des substances dangereuses dans l'air. La RFID ou d'autres technologies d'identification peuvent détecter si une personne est autorisée à rester dans une zone réglementée. Des capteurs peuvent surveiller la température des équipements industriels afin de vérifier s'ils fonctionnent en toute sécurité, que ce soit pour l'opérateur humain ou pour l'environnement. Bien entendu, plus les sources d'informations disponibles sont nombreuses, plus le système d'alerte sera efficace. Par ailleurs, une utilisation excessivement invasive de telles technologies peut donner lieu à des atteintes à la vie privée.

## **2.2. CLOUD COMPUTING**

La technologie en infonuagique (Cloud computing) est devenue une alternative réseau distribuée à hautes performances et à faible coût. Elle fournit des solutions de virtualisation matérielle, « la virtualisation : technologie qui extrait les détails du matériel physique et fournit des ressources virtualisées pour des applications de haut niveau » [23]. Il fournit un accès aux données à haute vitesse et une puissance de calcul complexe aux problèmes d'ingénierie à grande échelle via Internet. Internet est l'épine dorsale de la technologie cloud computing [24].

Avec la technologie cloud computing, une entreprise n'a pas besoin de posséder un matériel et des processeurs (GPU) coûteux pour résoudre des problèmes d'ingénierie complexes. Au lieu de cela, ils peuvent acheter un espace de stockage et une puissance de calcul haute performance sous forme de service, « matériel en tant que service (HaaS) et logiciel en tant que service (SaaS) ». Le cloud computing comprend le stockage matériel, les systèmes

d'exploitation, l'environnement d'exécution du programme, la base de données, le développement d'applications, les tests et la plateforme de déploiement. Les fournisseurs de services informatiques célèbres qui offrent des services en nuage sont entre autres : Microsoft, Google et Amazon [23].

### **2.3. BIG DATA**

L'analyse des données massives et l'intelligence d'affaires transforment les données en renseignements exploitables, en modèles prédictifs, en outils d'aide à la décision et en tableaux de bord.

Compte tenu des gigantesques volumes d'informations générés par les appareils connectés dans les chaînes de valeur du secteur de la fabrication, on peut dire que le Big Data se transforme progressivement en « Thing » Data. Les capteurs des appareils périphériques connectés peuvent théoriquement enregistrer leur comportement et leur statut en continu. Ces données d'événements sont alors filtrées, puis agrégées dans des référentiels Big Data gérés par l'intermédiaire de bases de données NoSQL. La recherche à partir du Big Data est ainsi à l'origine de solutions et de diagnostics proactifs, qui permettent d'anticiper et bien souvent d'éviter des incidents susceptibles d'engendrer d'onéreux processus de maintenance [8].

### **2.4. INDUSTRIE 4.0**

Les quatre grandes révolutions industrielles sont illustrées par une première révolution qui a été celle des machines à vapeur vers 1780. La deuxième fut celle de la production de masse avec l'électricité et le travail à la chaîne à partir de 1870. Ensuite vient la troisième révolution industrielle soit celle de l'automatisation, de l'électronique et des robots. De nos jours, plusieurs industries sont à cette étape dans leurs usines et d'autres sont toujours à la deuxième étape.



Aujourd'hui, nous sommes au stade de la quatrième grande révolution industrielle : celle de l'internet des machines industrielles et de l'usine intelligente. Les particularités de cette révolution très perturbatrice sont multiples. Il y'a par exemple : la rapidité à laquelle elle se déploie mondialement, le haut niveau de compétence exigée et la complexité accrue dans la gestion des données. L'intelligence artificielle devient de plus en plus la clé de voûte « incontournable » permettant d'utiliser de façon optimum la masse de données disponibles.

L'internet des objets, et plus particulièrement les technologies Machines à Machines liées au concept de l'Industrie 4.0, sont entraînent de prendre leur envol en Europe, avec comme leader, l'Allemagne. Selon une étude réalisée en 2015 par PwC& Strategy, le marché européen des machines connectées et intelligentes est appelé à exploser au cours des prochaines années. Au Québec et au Canada, l'utilisation de machines intelligentes ultra-efficaces permettra à l'industrie manufacturière de relever rapidement les défis tels que : l'augmentation de la productivité, la flexibilité de la production face aux besoins changeants des clients et surtout la résolution de la grave pénurie de main-d'œuvre annoncée par le vieillissement de la population active. Les cinq axes de technologies cités entourant l'usine intelligente ci-dessous amèneront rapidement les machines à apprendre, à prédire et à s'adapter à la production [25].

Il s'agit de :

- ✓ La robotique collaborative ;
- ✓ L'autoadaptation des systèmes de production ;
- ✓ La réalité virtuelle ou augmentée ;
- ✓ L'impression 3D.
- ✓ Et des capteurs intelligents et connectés ;

## **2.5. TYPES DE MAINTENANCE**

De nombreux types de maintenance coexistent afin de fournir aux industriels la possibilité d'optimiser leurs chaînes de production. Chaque entreprise définit donc ses besoins en matière de maintenance industrielle et met en place une stratégie de planification.

La maintenance est la combinaison de toutes les actions techniques et des actions administratives correspondantes, y compris les opérations de surveillance et de contrôle, destinées à maintenir ou à remettre une entité dans un état lui permettant d'accomplir une fonction requise [26]. Beaucoup de recherches sur la surveillance des roulements ont été faites vu l'importance de ces composantes dans l'industrie et la nécessité de faire un suivi fréquent et fiable. On distingue plusieurs types de mesures exploitées en industrie entre autres les analyses surs : les vibrations, l'huile et les mesures de température. Nous allons dans cette partie du mémoire présenter les principaux types de maintenance industrielle.

### **2.5.1. LA MAINTENANCE CORRECTIVE**

Comme l'indique son nom, la maintenance corrective vise à corriger les défauts de fonctionnement de tout équipement ou ligne de production dès leur apparition. Si la panne qui perturbe le processus de production n'a pas été anticipée par d'autres types de maintenance, comme la maintenance préventive, les techniciens chargés de la maintenance industrielle s'occupent de la résolution du problème dès leur survenue. Il s'agit ici d'une maintenance préventive non planifiée ou maintenance curative. Ce type d'intervention peut coûter cher aux entreprises.

### **2.5.2. LA MAINTENANCE PRÉVENTIVE**

Ce type de maintenance s'applique par les techniciens chargés de la maintenance industrielle en amont de toute apparition de panne ou de dysfonctionnement. Elle touche aux pièces détachées, aux composants ainsi qu'aux machines et aux équipements afin de réduire le risque d'échec ou de défaillance. La numérisation des entreprises industrielles a apporté de nombreuses solutions technologiques permettant aux techniciens d'effectuer, suivre et planifier la maintenance préventive efficacement.

### **2.5.3. LA MAINTENANCE PRÉDICTIVE**

L'émergence des solutions de traitement et d'analyse des données ainsi que de l'intelligence artificielle a permis aux industriels de planifier une maintenance prédictive ou prévisionnelle qui se base sur la prédiction des pannes et des dysfonctionnements. Ce type de maintenance industrielle permet aux entreprises d'anticiper les problèmes en planifiant les interventions nécessaires basées sur les prédictions. Elle permet ainsi de limiter les dépenses causées par les pannes inattendues et la perturbation de la production.

## **2.6. ANALYSE ET CLASSIFICATION DE DONNÉES**

En définition, l'analyse de données est un ensemble de méthodes descriptives ayant pour objectif de résumer et visualiser l'information pertinente contenue dans un grand tableau de données.

Les énormes progrès techniques récents en matière de puissance de calcul et de traitement, de capacité de stockage et d'interconnexion des différentes technologies de l'information ont créé des quantités sans précédent de données numériques. L'exploration de données est la science consistant à extraire des connaissances utiles de tels dépôts de données.

Elle est devenue un domaine en pleine croissance et interdisciplinaire en informatique. Les techniques d'extraction de données ont été largement appliquées aux problèmes de l'industrie, de la science et de l'ingénierie, ce qui prouve que l'extraction de données aura un impact profond sur notre société [27].

### **2.6.1. ANALYSE DE DONNÉE**

De nos jours, le volume des données collectées ne cesse d'augmenter exponentiellement et le critère du temps devient important. De ce fait, les techniques de classification automatique de données sont devenues très importantes pour faciliter l'exploitation de ces données. Plusieurs recherches et études ont été réalisées dans le domaine de la classification de forme pour permettre de mieux comprendre et d'exploiter les données surtout dans le cas où ces données sont de volume important et que les critères de sélection sont moins visibles.

Le processus de la classification de formes commence par une première étape d'acquisition et de prétraitement pour éliminer les différents bruits des données enregistrées. La deuxième étape consiste à définir les différentes caractéristiques à utiliser lors de la classification. Parmi ces caractéristiques, on trouve la moyenne, l'écart-type, la médiane. L'étape suivante consiste à développer le classificateur pour reconnaître les différentes classes et la dernière étape permet d'évaluer et valider le système de classification [28].

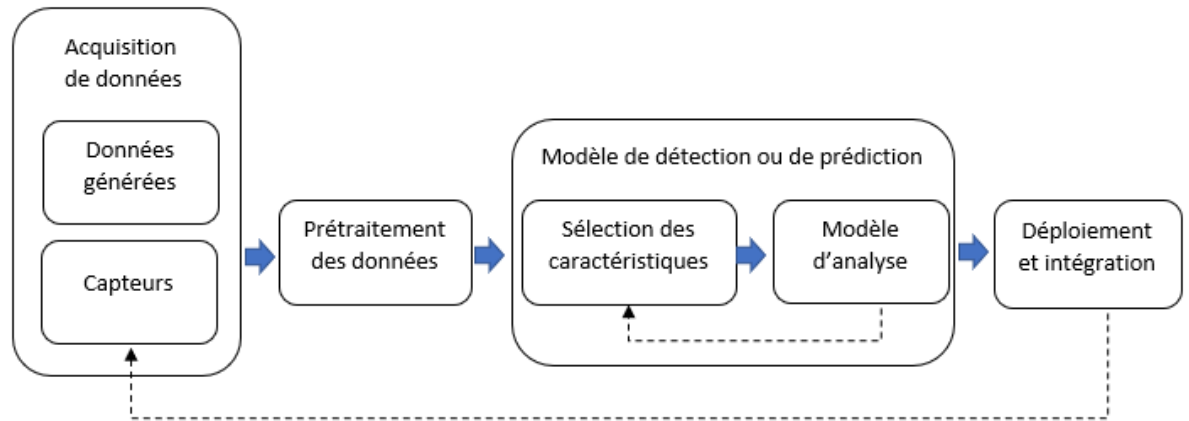


Figure 2 : Étapes d'analyse de donnée (classification) [29].

#### ➤ **ACQUISITION DES DONNÉES :**

Le but de l'acquisition de données est d'importer les données mesurées et générées par les capteurs et les organiser pour une utilisation dans les applications d'analyse de données.

#### ➤ **PRÉPARATION ET PRÉTRAITEMENT DES DONNÉES :**

Cette étape de l'analyse de données inclut le filtrage, le traitement dans le domaine temporel, le traitement dans le domaine fréquentiel et l'interpolation des signaux à étudier. Il est souvent nécessaire de nettoyer les données et les convertir dans un formulaire à partir duquel on pourra extraire des indicateurs de condition de prétraitement des données.

Généralement, le prétraitement consiste à nettoyer et transformer les données à analyser pour identifier et extraire des indicateurs de condition prometteuse. La maîtrise du type de système et du type de données à étudier permet de déterminer quelles méthodes de prétraitement seront utilisées.

➤ **TRANSFORMATION ET SÉLECTION DES CARACTÉRISTIQUES D'ANALYSE :**

Le but de cette étape est d'identifier les caractéristiques qui permettent de discriminer les différentes classes. Dans le cadre de l'analyse de l'activité physique, Oscar et al [30] ont recensé deux types de caractéristiques :

- ✓ Les caractéristiques basées sur la fréquence (caractéristiques fréquentielles).
- ✓ Les caractéristiques basées sur le temps (caractéristiques temporelles) telles que la moyenne, la médiane, l'écart-type, l'interquartile [31].

D'autres types de caractéristiques qui permettent une représentation globale peuvent être utilisées. En l'occurrence, l'analyse en composantes principales (ACP) qui est une méthode d'analyse de données très connue en statistique et dans les sciences expérimentales. Elle consiste à rechercher les directions de l'espace qui représentent le mieux les corrélations dans un ensemble de données. Ceci a pour objectifs de réduire la dimension des caractéristiques et de les visualiser afin d'analyser les corrélations entre ces données [28].

➤ **EXPLORATION ET CRÉATION DU MODÈLE DE CLASSIFICATION DES DONNÉES :**

La classification de formes est « l'opération qui consiste à grouper formellement des éléments de configuration, des incidents, des problèmes ou des changements, en fonction de leur type » (Office québécois de la langue française, 2005). En d'autres termes, la classification de formes consiste à affecter un objet à une classe correspondante en se basant sur ses caractéristiques. Les méthodes de classification de formes peuvent être réparties en deux types. La classification supervisée et la classification non supervisée [28].

## ➤ **INTERPRÉTATION ET ÉVALUATION :**

L'évaluation d'un système de classification se base généralement sur un partage de la base de données en trois domaines aussi appelés sous-bases :

- ✓ D'apprentissage utilisé pour la conception du système de classification ;
- ✓ De test utilisé pour sélectionner l'architecture du système de classification ;
- ✓ De validation utilisée pour valider la pertinence de l'architecture choisie.

La distinction entre les deux premières bases est nécessaire pour ne pas fausser les tests de classification en testant sur des échantillons ayant servi pour la conception. La troisième sous-base permet d'obtenir une validation indépendante des critères de sélection. Dans certains cas, la taille de la base de données ne permet pas d'avoir un partage en trois sous-bases ; dans ce cas, on utilise uniquement un partage en deux : une sous-base d'apprentissage et une sous-base de tests [28].

## ➤ **DÉPLOIEMENT :**

Le déploiement ou intégration d'un algorithme d'analyse de données est généralement la dernière étape du flux de travail pour la conception d'un système de classification. Elle représente la mise en exploitation du système.

### 2.6.2. CLASSIFICATION DE DONNÉES

La première étape d'un projet d'apprentissage automatique est la récupération et l'importation des données. Malheureusement, dans des projets réels, nous obtenons régulièrement des fichiers incomplets, comportant des erreurs. Il est donc crucial, avant même de travailler sur des visualisations ou des algorithmes, de s'assurer d'obtenir des données correctes, et bien étiquetées. L'apprentissage automatique comporte généralement deux phases. La première consiste à estimer un modèle à partir de données, appelées observations, qui sont disponibles et en nombre fini, lors de la phase de conception du système. La seconde phase correspond à la mise en production : le modèle étant déterminé, de nouvelles données peuvent alors être soumises afin d'obtenir le résultat correspondant à la tâche souhaitée.

Selon les informations disponibles durant la phase d'entraînement, l'apprentissage est qualifié de différentes manières. Si les données sont étiquetées (c'est-à-dire que la réponse à la tâche est connue pour ces données), il s'agit d'un apprentissage supervisé. On parle de classification si les étiquettes sont discrètes, ou de régression si elles sont continues. Si le modèle est appris de manière incrémentale en fonction d'une récompense reçue par le programme pour chacune des actions entreprises, on parle d'apprentissages par renforcement. Dans le cas le plus général, sans étiquette, on cherche à déterminer la structure sous-jacente des données (qui peuvent être une densité de probabilité) et il s'agit alors d'apprentissage non supervisé.

Les différences entre l'analyse de regroupement (ou clustering) et l'analyse discriminante ne concernent que la session d'apprentissage de l'algorithme. Une fois que les paramètres sont déterminés et que nous commençons à utiliser le modèle, les deux modèles ont le même usage pour classer un objet dans un certain nombre de catégories.



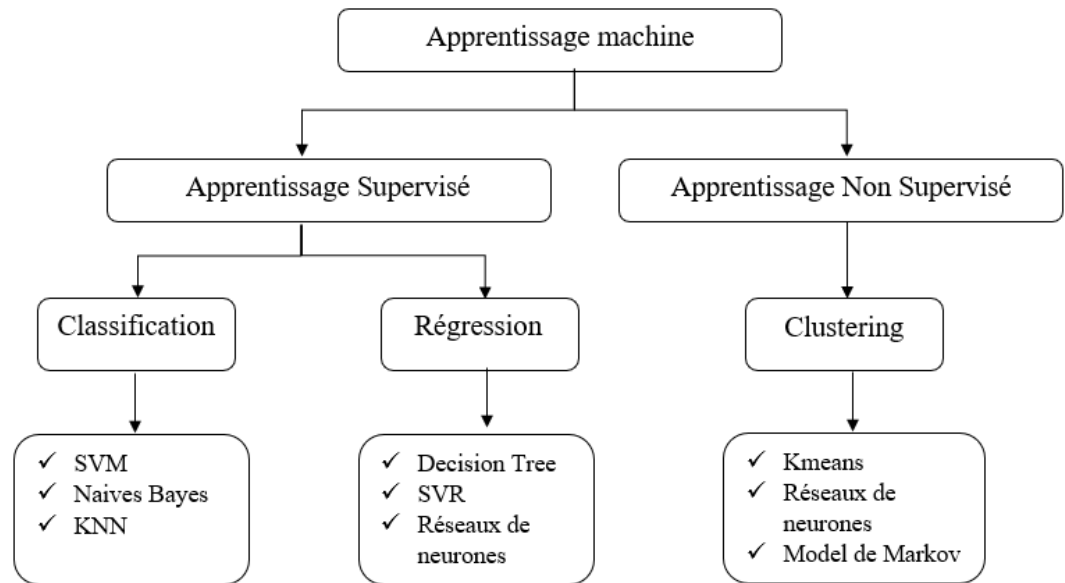


Figure 3 : Taxonomie des modèles d'apprentissage automatique [32].

**Tableau 1 : Comparaison des types d'analyses de classification.**

	Analyse par classification	Analyse discriminante
Nom	Apprentissage non supervisé	Apprentissage supervisé
Période d'apprentissage	La catégorie d'objet est inconnue	La catégorie d'objet est connue
But de l'apprentissage	Connaître la catégorie d'objet inconnue	Connaître la règle de classement
Après l'apprentissage	Pour classer un objet dans un certain nombre de catégories	Pour classer un objet dans un certain nombre de catégories

### **2.6.2.1. APPRENTISSAGE SUPERVISÉ**

#### **➤ CLASSIFICATION**

Dans l'apprentissage automatique et les statistiques, la classification est le problème qui consiste à identifier à quel groupe de catégories (sous-populations) une nouvelle observation appartient, à partir d'un ensemble d'apprentissages contenant des données (ou instances) dont l'appartenance à une catégorie est connue. Par exemple, attribuer un courrier électronique donné à la classe "spam" ou "non-spam" et attribuer un diagnostic à un patient donné en fonction des caractéristiques observées du patient (sexe, pression artérielle, présence ou non de certains symptômes, etc.). La classification est considérée comme un exemple d'apprentissage supervisé et de reconnaissance de formes, c'est-à-dire un apprentissage dans lequel un ensemble de formations d'observations correctement identifiées est disponible. La procédure non supervisée correspondante est connue sous le nom de clustering. Elle consiste à regrouper des données en catégories en fonction d'une mesure de la similarité inhérente ou de la distance [33].

#### **➤ RÉGRESSION**

L'analyse de régression est largement utilisée pour la prévision, pour comprendre quelles variables indépendantes sont liées à la variable dépendante et pour explorer les formes de ces relations. Dans des circonstances restreintes, une analyse de régression peut être utilisée pour déduire des relations de cause à effet entre les variables indépendantes et dépendantes.

#### **2.6.2.2. APPRENTISSAGE NON SUPERVISÉ**

En clustering, la catégorie de l'objet est inconnue. Cependant, nous connaissons la règle à classer (généralement basée sur la distance) et nous connaissons également les caractéristiques (variables indépendantes) pouvant décrire la classification de l'objet. Il n'y a pas d'exemple de formation pour examiner si la classification est correcte ou non. Ainsi, les objets sont assignés à des groupes simplement basés sur la règle donnée.

### **2.7. REVUE DE LA LITTÉRATURE**

L'analyse des vibrations est l'une des techniques les plus utilisées pour la maintenance prédictive dans les machines tournantes à grande vitesse. En utilisant les informations contenues dans les signaux de vibration, des études ont porté sur la conception d'un système de détection et de diagnostic des défaillances de composants mécaniques. Étant donné que la collecte de données sur les défaillances précédentes est irréalisable dans des scénarios réels, une méthode utilisant le paradigme one-class-v-SVM est utilisée [34].

D'autres recherches dans nos lectures ont utilisé le paramètre de dégradation de l'état normal à l'état de panne du système, qui est nécessaire en tant qu'objet de surveillance de l'état de fonctionnement dans la maintenance. Des articles proposent un modèle de Markov caché pour évaluer la dégradation de l'état de la machine. Une analyse dynamique en composantes principales et la distance de Mahalanobis sont utilisées pour extraire les caractéristiques efficaces des signaux de vibration permettant de quantifier l'état de santé de la machine [34-36]; puis d'utiliser ces données pour estimer la durée de vie utile restante (RUL) des machines [35].

Des articles proposent d'appliquer l'entropie d'échantillons de paquets d'ondelettes dans la prévision de la tendance des défaillances des machines roulant. Cette méthode permettrait de mieux prévoir l'état de fonctionnement des roulements.

Ainsi, la méthode d'entropie des échantillons de paquets d'ondelettes peut être utilisée comme un bon outil pour la surveillance et la prévision des roulements [37].

Nous avons également découvert dans nos recherches qu'un logiciel appelé cbmLAD, en cours de développement à l'École Polytechnique de Montréal utilise une approche d'exploration de données d'apprentissage supervisé appelée analyse logique des données (LAD) [38].

## **CHAPITRE 3**

### **MÉTHODOLOGIE**

En utilisant des capteurs pour déterminer à quel moment la vérification des équipements est nécessaire, il est possible de prévenir les pannes et de réduire les coûts de la maintenance de routine. Grâce à des capteurs intégrés et connectés à internet, le contrôle des équipements de production s'effectue à distance et en temps réel. Dans ce cas, les recommandations sont envoyées aux services de l'exploitation et de la maintenance afin de remédier aux problèmes avant même qu'ils ne surviennent. Cette méthode permet de réduire les coûts d'exploitation et les coûts d'investissement en favorisant des réparations et une maintenance proactive des équipements en vue d'améliorer l'utilisation de la capacité et de la productivité [1].

#### **3.1. INTRODUCTION**

La méthode développée dans ce projet est testée sur des données de référence extraites du « référentiel de données pronostiques de la NASA » et relatives à plusieurs expériences de défaillances des roulements effectuées dans différentes conditions de fonctionnement. En outre, la méthode est comparée aux pronostics traditionnels de caractéristiques temporelles et fréquentielles dont les résultats de simulation sont présentés à la fin du document.

#### **3.2. PROTOCOLE D'ACQUISITION DES SIGNAUX**

Les données ont été générées par le centre National Science Foundation (NSF) Industry–University Cooperative Research Centers (IUCRC) pour les systèmes de maintenance intelligents (IMS - [www.imscenter.net](http://www.imscenter.net)) avec le soutien de Rexnord Corporation à Milwaukee, Wisconsin États-Unis.

Quatre roulements ont été installés sur un arbre. La vitesse de rotation a été maintenue constante à 2000 tr/min par un courant alternatif. Le moteur est couplé à l'arbre via des courroies de frottement. Une charge radiale de 6000 lb est appliquée sur l'arbre et le roulement par un mécanisme à ressort. Tous les roulements sont lubrifiés à force. Les roulements à double rangée Rexnord ZA-2115 ont été installés sur l'arbre comme indiqué à la figure 4. Une Carte PCB 353B33 avec des accéléromètres Quartz ICP à haute sensibilité a été installée sur le corps de palier avec un accéléromètre pour chaque roulement. La figure 4 indique également la position des capteurs. Toutes les défaillances sont survenues après avoir dépassé la durée de vie prévue du roulement qui est de plus de 100 millions de révolutions [9].

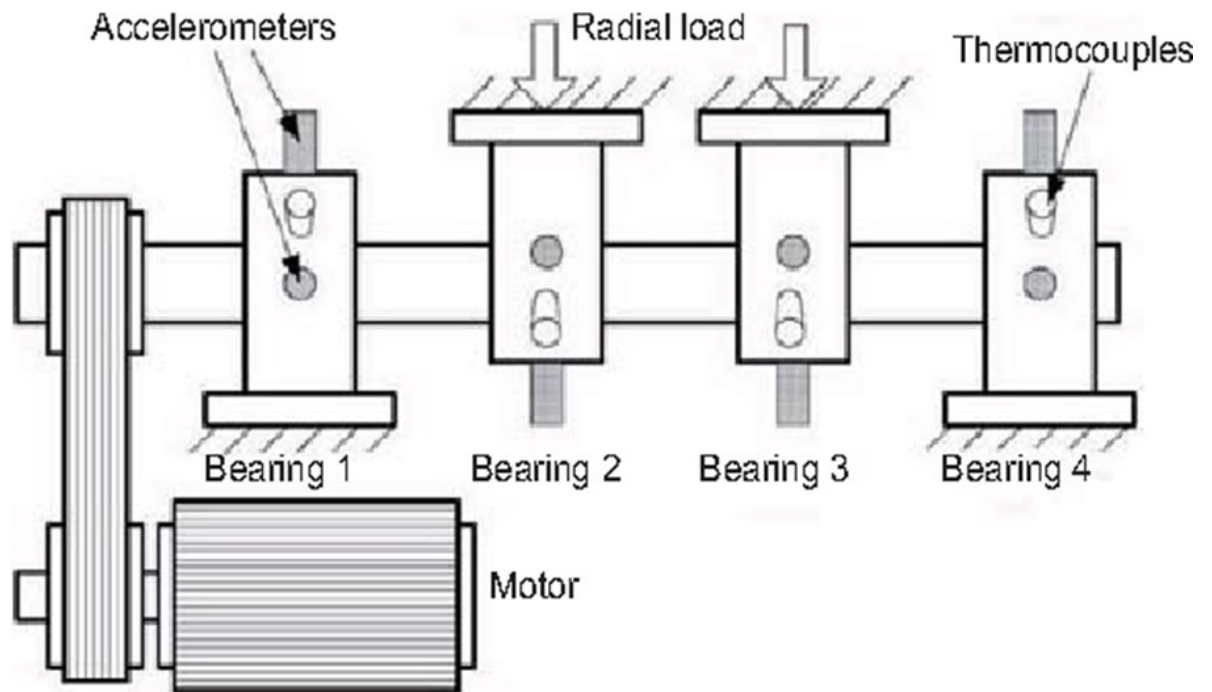


Figure 4 : Banc d'essai et illustrateur de placement des capteurs [9].

### 3.3. PRÉTRAITEMENT DES SIGNAUX

Pour le prétraitement de nos signaux, nous avons utilisé la boîte à outils Signal Processing Toolbox du logiciel MATLAB qui permet grâce à ses fonctions de détecter les valeurs aberrantes et de les corriger. Il permet également de travailler avec des signaux échantillonnés irrégulièrement, et de les préparer pour une analyse plus poussée [39].

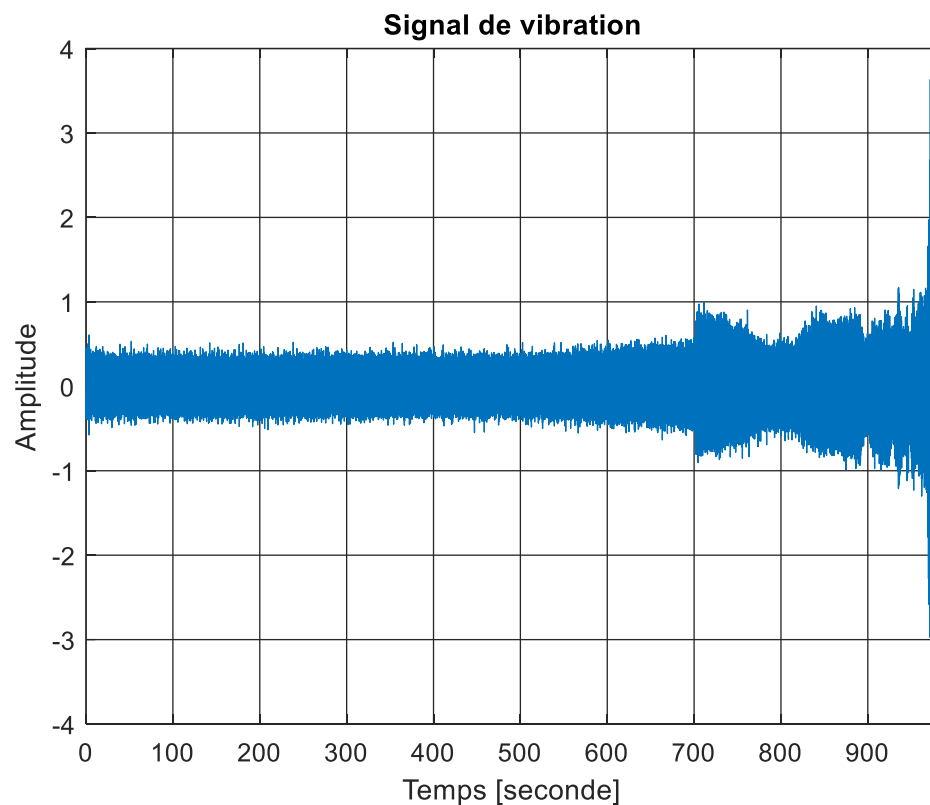


Figure 5 : Signal de vibration de la machine enregistré jusqu'à la défaillance.

Nous avons pour les besoins de notre projet découpé notre signal en trois groupes de données en fonction du niveau de dysfonctionnement observé durant le cycle de vie de la machine. Pour la conception du système de surveillance de la machine, nous avons procédé au filtrage de notre signal en appliquant un filtre médian (MED) aux signaux de vibration pour conserver les informations utiles en éliminant les composantes de bruit à haute fréquence.

Étant donné que les spectrogrammes  $P(t, \omega)$  des données des roulements sains et défectueux sont différentes, des caractéristiques représentatives peuvent être extraites des spectrogrammes et utilisées pour calculer avec précision la fréquence de crête moyenne comme indicateur de l'état du système [40] :

$$PeakFreq(t) = \operatorname{argmax}_{\omega} \{P(t, \omega)\} \quad (3.1)$$

La fréquence de crête moyenne est la moyenne des fréquences de crête définies ci-dessus :

$$MeanPeakFreq = \frac{1}{T} \int_0^T PeakFreq(t) dt \quad (3.2)$$

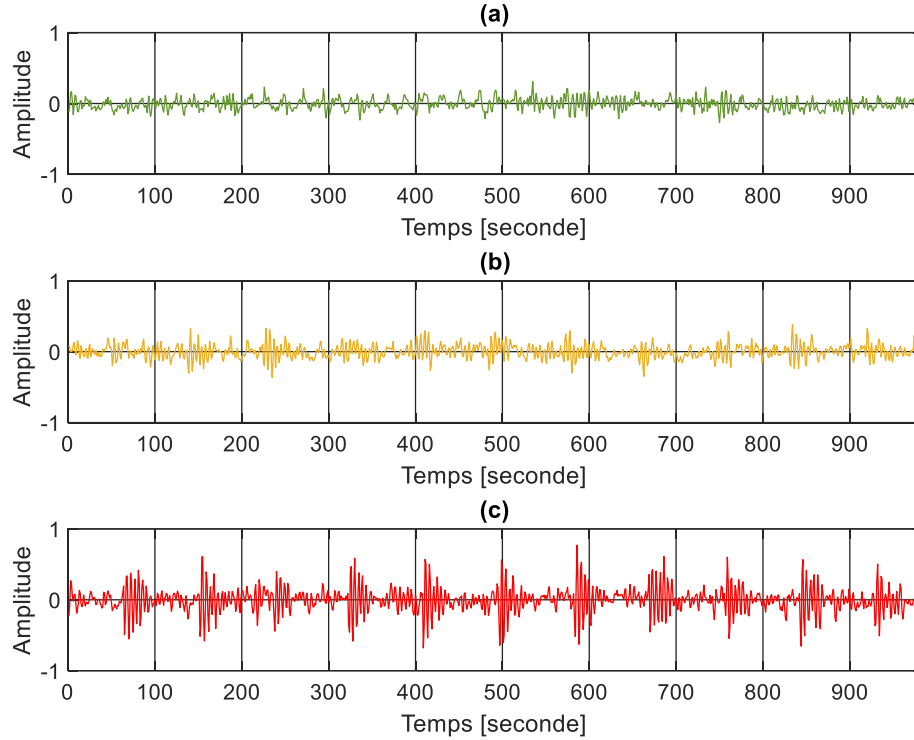


Figure 6 : Découpage du signal en trois groupes selon la présence ou non de défauts.  
(a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.



### 3.4. SÉLECTION DES CARACTÉRISTIQUES

Le logiciel MATLAB fournit des fonctions qui permettent d'explorer et d'extraire des modèles dans les signaux. Plus précisément, localiser les pics de signal et déterminer leur hauteur, leur largeur et trouver des points de changement dans les signaux [39].

Les caractéristiques du signal fournissent des métriques statistiques générales basées sur le signal qui peuvent être appliquées à tout type de signal, y compris un signal de vibration. Les modifications apportées à ces fonctionnalités peuvent indiquer des modifications de l'état de santé du système.

#### 3.4.1. CARACTÉRISTIQUES TEMPORELLES

L'analyse du domaine temporel présente les inconvénients d'une faible sensibilité et d'une faible précision, mais ses calculs simples et son traitement direct du signal contribuent à raccourcir le temps de traitement. La méthode simple du domaine temporel ne convient pas pour un diagnostic efficace des défauts, mais elle est bien meilleure lorsqu'elle est combinée avec d'autres approches, par exemple, les méthodes de classification, le réseau neuronal et l'intelligence artificielle. C'est-à-dire en utilisant des mesures statistiques pour obtenir des fonctionnalités utiles, puis en les classant par exemple avec un algorithme d'arbre de décision.

Les caractéristiques temporelles capables d'identifier l'emplacement précis des défauts sont les suivantes :

➤ **CARRE DE LA MOYENNE QUADRATIQUE (ROOT MEAN SQUARE (RMS)) :**

Qui indique l'état général aux derniers stades de dégradation. RMS est sensible à la charge et aux changements de vitesse.

- **FACTEUR DE CRÊTE** : Qui indique un rapport entre la valeur de crête du signal et la valeur RMS indiquant les signes précoces de dommages, en particulier lorsque les signaux de vibration présentent des caractéristiques impulsives.
- **KURTOSIS** : Indique les pics principaux de la distribution d'amplitude des signaux de vibration. C'est un paramètre calculé statistiquement à partir de la forme d'onde temporelle qui informe de la quantité de pics présents dans la forme d'onde.

**Tableau 2 : Caractéristiques temporelles**

Caractéristiques	Expression mathématique
Moyenne	$\frac{1}{N} \sum_{i=1}^N  x_t(i)  \quad (3.3)$
Root Mean Square (RMS)	$\sqrt{\frac{1}{N} \sum_{i=1}^N x_t(i)^2} \quad (3.4)$
Valeur pic (PEAK)	$\text{Max}(x_t) \quad (3.5)$
Facteur de crête	$\frac{VP(x_t)}{RMS(x_t)} \quad (3.6)$
Coefficient d'asymétrie (skewness en anglais)	$\frac{1}{N} \sum_{i=1}^N \left[ \frac{x_t(i) - \bar{x}_t}{\sigma_s} \right]^3 \quad (3.7)$
Kurtosis	$\frac{1}{N} \sum_{i=1}^N \left[ \frac{x_t(i) - \bar{x}_t}{\sigma_s} \right]^4 \quad (3.8)$

Où :

$x_t$  = Signal brut,

$N$  = Le nombre de points enregistrés à l'instant (t)

$\sigma_s$  = L'écart-type du signal.

Les caractéristiques issues du Kurtosis sont obtenues en considérant un moment d'ordre 4 du signal et se calculent selon la formule (3.8) présentée dans le tableau 2. Dans le même ordre d'idée, le Skewness est un indicateur d'ordre 3 mesurant l'énergie par rapport à la moyenne.

Beaucoup de recherches ont fait ressortir l'efficacité du Kurtosis, pour la détection des défauts de roulements. Il s'agit d'un indicateur représentant les impulsions dans un signal. Certains articles de recherche [41] ont démontré l'utilité de l'utilisation du Kurtosis pour la détection des défauts de roulements par rapport aux autres indicateurs conventionnels tels que le PEAK ou le RMS. Cet indicateur permet d'avoir une idée sur la santé du roulement sans pour autant savoir beaucoup de choses sur l'historique de la machine. En effet, une valeur supérieure à 3 signifie que le roulement est défectueux [42], sauf que le Kurtosis peut être plus sensible qu'il ne le faut et peut donner des alarmes prématurément [43].

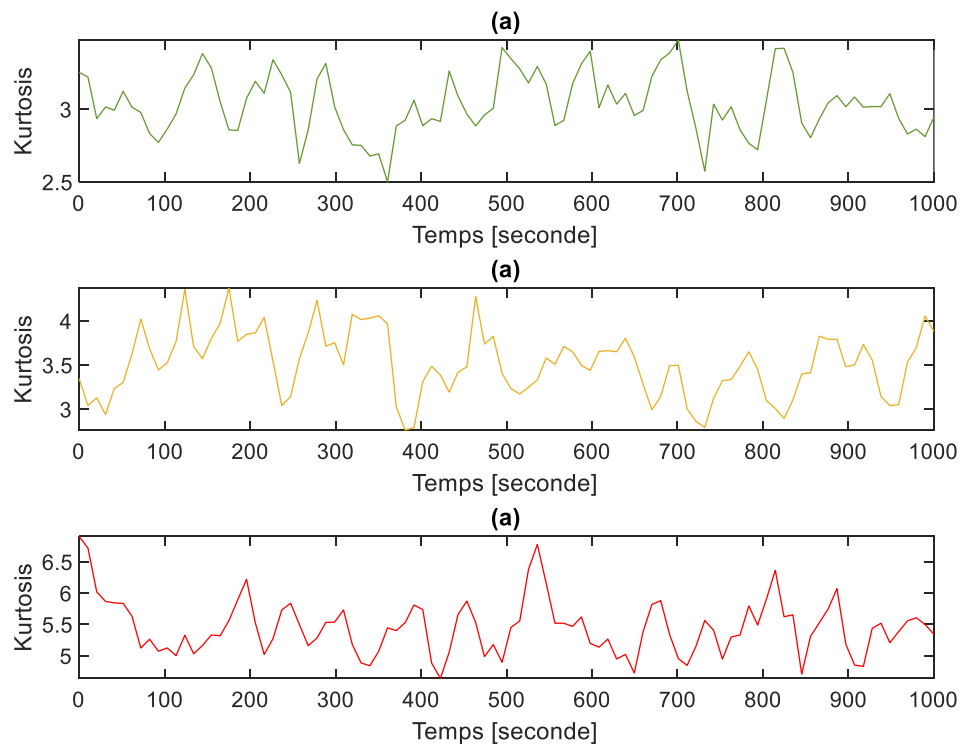


Figure 7 : Spectre de kurtosis dans le domaine temporel des signaux. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.

### 3.4.2. CARACTÉRISTIQUES FRÉQUENTIELLES

L'analyse spectrale est déterminante pour diagnostiquer l'endommagement d'un roulement, ce qui nous permet, par comparaison de l'évolution des amplitudes spectrales, de prédire le degré de détérioration du roulement et de planifier son remplacement. Dans cette partie, nous approfondirons l'utilisation de ces techniques.

La détérioration des éléments générera une ou plusieurs fréquences caractéristiques défaillantes dans les spectres de fréquences qui nous permettront une identification rapide et facile. Les quatre fréquences de défaillance de roulement possibles sont [44] :

- ✓ **BPFO** (Ball Pass Frequency Outer) ou fréquence d'échec de course externe : Qui correspond physiquement au nombre de billes ou de rouleaux qui passent par un point donné de la course extérieure à chaque fois que l'arbre effectue un tour complet.
- ✓ **BPFI** (Ball Pass Frequency Inner) ou fréquence d'échec de la course interne : Qui correspond physiquement au nombre de billes ou de rouleaux qui passent par un point donné de la piste intérieure chaque fois que l'arbre effectue un tour complet.
- ✓ **BSF** (Ball Spin Frequency) ou fréquence de défaillance des éléments roulants : Qui correspond physiquement au nombre de tours qu'une bille ou un rouleau de roulement effectue chaque fois que l'arbre effectue un tour complet.
- ✓ **FTF** (Fundamental Train Frequency) ou fréquence de défaillance de cage : Qui correspond physiquement au nombre de tours que fait la cage de roulement à chaque fois que l'arbre effectue un tour complet.

**Tableau 3 : Caractéristiques fréquentielles**

Caractéristiques	Expression mathématique
BPFO	$\frac{Nb}{2} \left( 1 - \frac{Bd}{Pd} \cos \theta \right) \times \omega \quad (3.9)$
BPFI	$\frac{Nb}{2} \left( 1 + \frac{Bd}{Pd} \cos \theta \right) \times \omega \quad (3.10)$
BSF	$\frac{Pd}{2Bd} \left[ 1 - \left( \frac{Bd}{Pd} \cos \theta \right)^2 \right] \times \omega \quad (3.11)$
FTF	$\frac{\omega}{2} \left( 1 \pm \frac{Bd \cos \theta}{Pd} \right) \quad (3.12)$

Où :

$Nb$  = Nombre de billes

$Bd$  = Diamètre des billes

$Pd$  = Pas diamétral

$\theta$  = Angle de contact

$\omega$  = La vitesse de rotation de la partie tournante

La technique la plus répandue pour extraire ce type de caractéristiques est la transformée de Fourier discrète et rapide (Fast Fourier Transform : FFT). La transformée de Fourier discrète  $X(k)$  d'un signal échantillonné  $x(n)$  composé de  $N$  échantillons est donnée par l'expression suivante :

$$X(k) = \sum_{n=0}^{N-1} x(n). e^{-2i\pi \frac{k}{N}} \quad (3.13)$$

Où :  $0 \leq k \leq N$

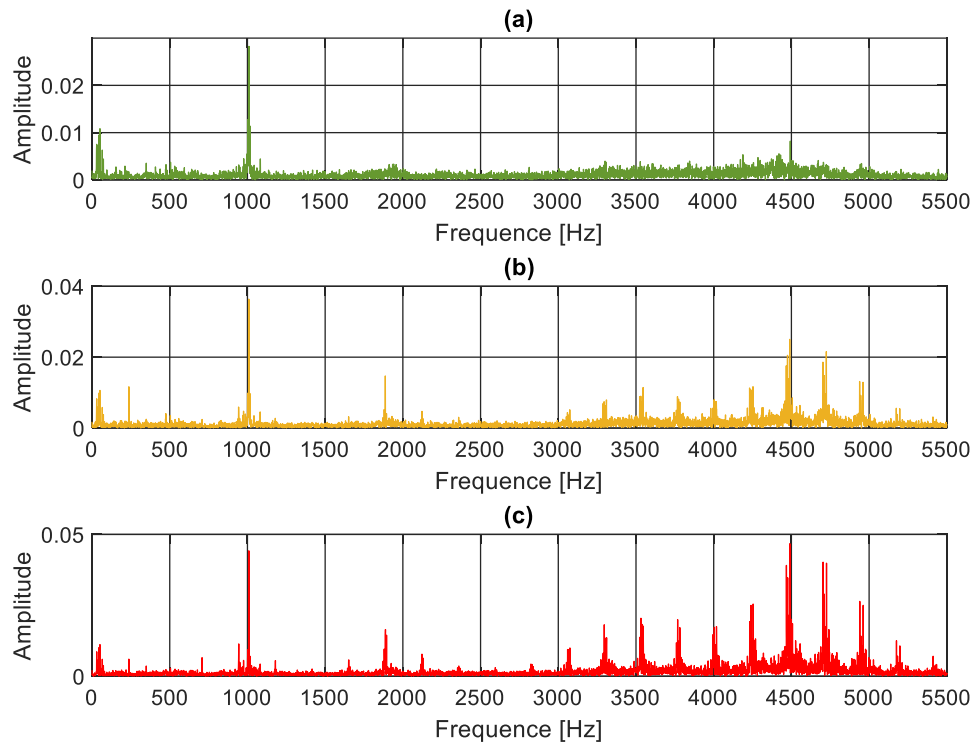


Figure 8 : Spectre d'amplitude des signaux obtenus de la FFT. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.

### **3.4.3. CARACTÉRISTIQUES TEMPS-FRÉQUENCE**

L'analyse temps-fréquence fournit les informations de distribution conjointe du domaine temporel et fréquentiel, qui illustrent clairement que la fréquence du signal variant en fonction du temps. Pour caractériser l'intensité énergétique d'un signal à différents moments et fréquences, diverses approches ont été proposées, telles que la transformée de Fourier à court terme [45], la distribution de Wigner-Ville [46] et la transformée en ondelettes continues [47]. Cependant, le calcul de ces méthodes prend trop de temps pour que le processus de classification devienne plus compliqué. Bien que de nombreuses recherches aient été menées dans le domaine du diagnostic des défauts, il est évident que très peu de littératures ont signalé l'amélioration des algorithmes pour reconnaître efficacement les défauts de taille micro [48].

### **3.4.4. LES LIMITES DES APPROCHES TEMPS-FRÉQUENCE**

La plupart des indicateurs temporels ne sont réactifs à de basses vitesses qu'à un stade assez avancé de la dégradation du roulement. Mais la limitation réelle derrière l'approche temps-fréquence est le temps de calcul et les ressources que consomme cette méthode. Étant donné que l'objectif principal est de mettre en place un système de surveillance permanente pour des machines tournantes, le temps de calcul est un paramètre très important à considérer.

## **3.5. SYSTÈME DE SURVEILLANCE D'ÉTAT DE LA MACHINE**

Dans cette partie du projet, nous avons mis en place un système largement inspiré de [40] qui explique comment extraire des caractéristiques des données mesurées pour effectuer une surveillance et des pronostics. Sur la base des fonctionnalités extraites, des modèles dynamiques sont générés, validés et utilisés pour prévoir le temps des pannes afin que des actions puissent être entreprises avant que les pannes réelles ne se produisent. Donc l'idée est d'implémenter des seuils de références dans cette approche temps-fréquence, pour que le système de surveillance soit capable de donner l'état des machines en adoptant cette approche.

### 3.6. LES ALGORITHMES DE CLASSIFICATION

Après la description de l'utilité des caractéristiques à la section 3.4. Nous constatons qu'elles servent à construire un vecteur pour la classification des défauts à l'aide d'algorithme de classification. Ces algorithmes permettent, dans la grande majorité des cas, d'obtenir de très bon taux de reconnaissance. En effet, le principe de base repose sur un ensemble de données du problème qui servent à entraîner l'algorithme de classification ce qui permet alors d'exploiter des outils puissants dans le domaine de l'apprentissage automatique. Au vu de notre objectif de recherche qu'est la détection préventive des défauts de la machine tournante, ces algorithmes semblent être la solution à déployer selon différentes stratégies d'approches.

L'apprentissage supervisé est un type d'apprentissage automatique qui utilise un jeu de données connu pour effectuer des prédictions. L'ensemble de données d'apprentissage comprend les données d'entrée et les valeurs de réponse. Les algorithmes d'apprentissage supervisé cherchent à créer un modèle capable de prédire les valeurs de réponse d'un nouvel ensemble de données. L'utilisation de jeux de données d'apprentissage plus volumineux et l'optimisation des hyperparamètres du modèle peuvent souvent augmenter le pouvoir prédictif du modèle et lui permettre de bien généraliser pour les nouveaux jeux de données. Un ensemble de données de test est souvent utilisé pour valider le modèle.

Les modèles de classification sont utilisés pour affecter des éléments à un groupe ou une classe discrète en fonction d'un ensemble spécifique de fonctionnalités. Chaque modèle a ses propres forces et faiblesses dans un scénario donné. Le choix d'un modèle de classification des données est étroitement lié à l'analyse de rentabilisation et à une solide compréhension de ce qu'on essaye d'accomplir. Le choix de la bonne approche nécessite souvent des essais et des erreurs pour atteindre le bon équilibre entre complexité, performances et précision. Les sections suivantes décrivent certains des modèles courants qui sont utiles à connaître.



Dans notre projet, nous avons utilisé la boîte à outils (Statistics and Machine Learning Toolbox) du logiciel MATLAB pour faire la classification de nos signaux qui peuvent être séparés en "classes" spécifiques. Les tests ont porté sur les algorithmes de classification suivante :

### 3.6.1. ARBRES DE DÉCISION

Un arbre de décision est un enchaînement hiérarchique de règles logiques ou de production construites de manière automatique à partir d'un ensemble de données. La construction de l'arbre de décision consiste à utiliser les descripteurs, pour subdiviser progressivement l'ensemble en sous-ensembles de plus en plus fins. L'algorithme estime la probabilité qu'une observation se trouve dans le nœud  $t$  en utilisant la formule :

$$P(T) = \sum_{j \in T} w_j \quad (3.14)$$

Où :

$w_j$  = Poids des observations,

$T$  = Ensemble de tous les indices d'observation du nœud  $t$

$n$  = Égale au nombre d'observations.

Les arbres de décision sont des algorithmes de classification de données reposant sur un apprentissage supervisé. Comme son nom l'indique, la structure de ces algorithmes ressemble à des arbres constitués de nœuds, de branches et de feuilles. La construction de ces arbres est réalisée à l'aide d'une base de données brute (vecteur des caractéristiques et classes) et de lois qui permettent de déterminer les variables discriminantes pour la classification efficace des futures données. Chacun des nœuds constituant l'arbre représente une règle de classification préalablement déterminée de manière récursive.

Dans les domaines de l'informatique décisionnelle et du forage de données, ce type d'algorithmes est très utilisé, car ils permettent de construire automatiquement des classifications et leur vitesse d'exécution surpasse tous les autres types des algorithmes. Cependant, ils ne permettent pas d'obtenir de bons résultats quand les données sont complexes.

Pour voir comment un arbre de décision prédit une réponse, il faut suivre les décisions dans l'arbre du nœud racine (début) jusqu'au nœud feuille qui contient la réponse. Les arbres de classification donnent des réponses nominales, tel que vrai ou faux et les arbres de régression donnent des réponses numériques. Les arbres de décision sont relativement rapides et faciles à suivre dans leur résolution ; vous pouvez voir une représentation complète du chemin parcouru de la racine à la feuille. Cela est particulièrement utile si vous devez partager les résultats avec des personnes intéressées par la manière dont une conclusion a été tirée.

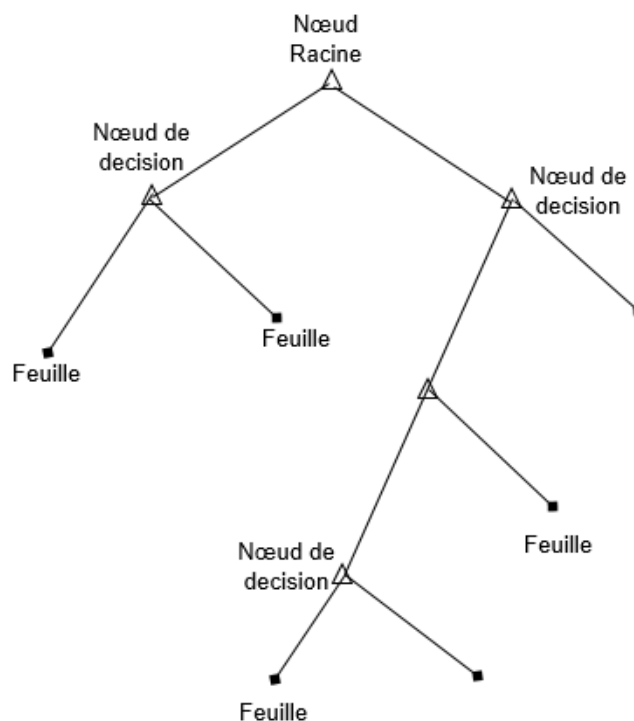


Figure 9 : Exemple d'illustration d'un arbre de décision.

### 3.6.2. MACHINES A VECTEURS DE SUPPORT (SVM)

Les machines à vecteurs de support (SVM) ont été développées à l'origine par Vapnik et ses collègues en 1992 sur la base de la théorie de l'apprentissage statistique de Vapnik & Chervonenkis dans les années 1960. Le SVM a été appliqué avec succès dans de nombreuses applications, notamment la reconnaissance manuscrite, la prédiction de séries chronologiques, la reconnaissance vocale, le problème de séquence protéique, le diagnostic du cancer du sein et bien d'autres [49].

Les machines à vecteur de support (SVM) sont utilisées lorsque les données ont exactement deux classes. L'algorithme SVM classe les données en trouvant le meilleur hyperplan qui sépare tous les points de données d'une classe de ceux de l'autre classe (le meilleur hyperplan pour un SVM est celui avec la plus grande marge entre les deux classes). L'algorithme SVM peut aussi être utilisé avec plus de deux classes, auquel cas le modèle créera un ensemble de sous-ensembles de classification binaire.

Il y a quelques avantages importants à utiliser l'algorithme SVM. Tout d'abord, il est extrêmement précis et n'a pas tendance à suradapter les données. Deuxièmement, les machines à vecteurs de support linéaire sont relativement faciles à interpréter. Parce que les modèles SVM sont très rapides, une fois que votre modèle a été formé, vous pouvez supprimer les données de formation si vous avez une capacité de mémoire disponible limitée. Il a également tendance à très bien gérer les classifications complexes et non linéaires en utilisant une technique appelée « astuce du noyau ». Cependant, les algorithmes SVM doivent être formés et réglés à l'avance, vous devez donc investir du temps dans le modèle avant de pouvoir commencer à l'utiliser. De plus, sa vitesse est fortement affectée si vous utilisez le modèle avec plus de deux classes.

L'algorithme SVM est un classificateur dit linéaire, ça veut dire que, dans le cas parfait, les données doivent être linéairement séparables. Il permet de trouver le meilleur séparateur (ligne, plan ou hyperplan) qui sépare le mieux les deux classes.

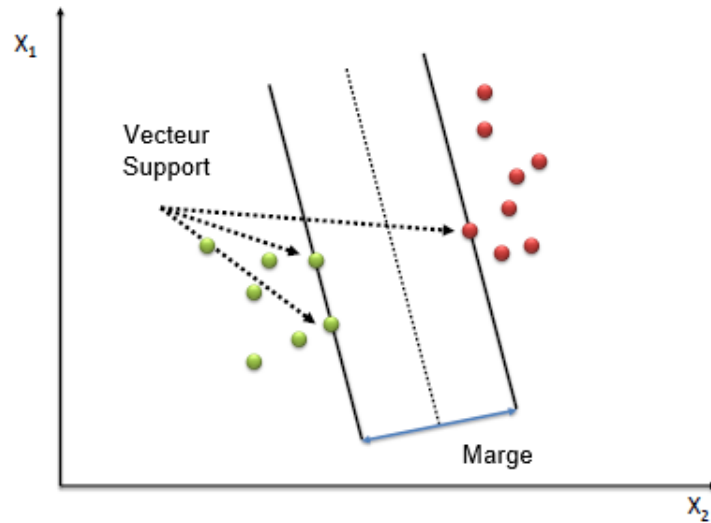


Figure 10 : Principe de l'algorithme SVM.

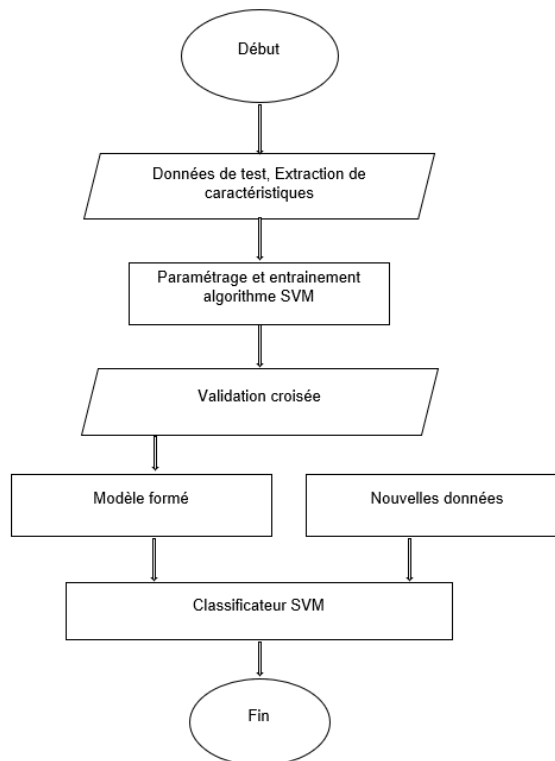


Figure 11 : Organigramme de l'algorithme SVM.

### 3.6.3. LES k-PLUS PROCHES VOISINS (k-NN)

L'algorithme des k-plus proches voisins (k-NN) se base sur les données en entier. En effet, pour une observation, qui ne fait pas partie des données, qu'on souhaite prédire, l'algorithme va chercher les k instances les plus proches de notre observation et choisir pour chaque observation la classe majoritaire parmi ses k plus proches voisins.

La méthode k-NN est une technique d'apprentissage supervisé, et est considérée comme l'une des plus simples dans le domaine de la classification. Elle permet de classer une nouvelle observation (vecteur de caractéristiques extraites) en calculant la distance avec les données d'entraînement, et de prendre les k plus proches voisins (en termes de distance). Puis, observé la classe qui est majoritairement représentée parmi les k-plus proches voisins et d'assigner cette classe à la nouvelle observation.

Bien que le temps d'apprentissage de l'algorithme k-NN soit court, le temps de requête réel (et l'espace de stockage) peut être plus long que celui des autres modèles. Cela est particulièrement vrai lorsque le nombre de points de données augmente, car toutes les données d'entraînement doivent être conservées, mais pas seulement l'algorithme.

Le plus grand inconvénient de cette méthode est qu'elle peut être erronée par des attributs non pertinents qui masquent des attributs importants. Il existe des moyens de corriger ce problème, par exemple en appliquant des pondérations aux données. Comme nous l'avons détaillé ci-dessus, l'algorithme k-NN calcule la distance entre les points de données. Pour cela, nous utilisons la formule de la distance euclidienne :

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.15)$$

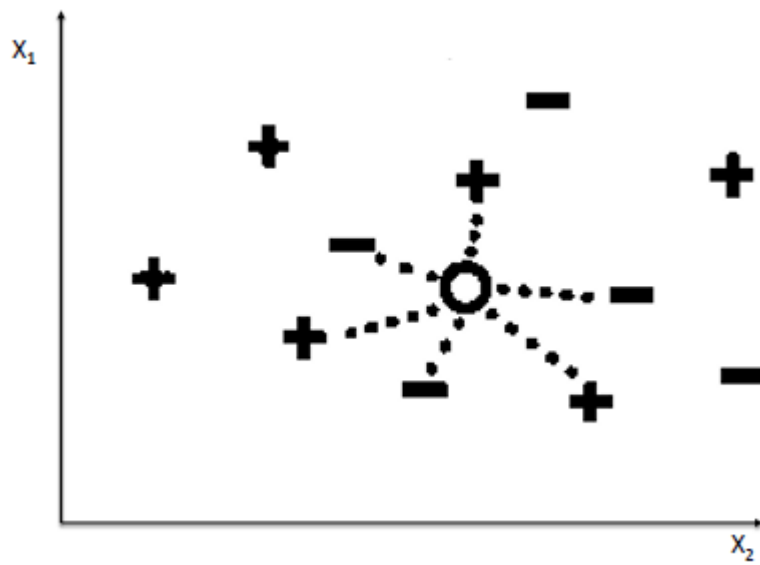


Figure 12 : Principe de l'algorithme k-NN.

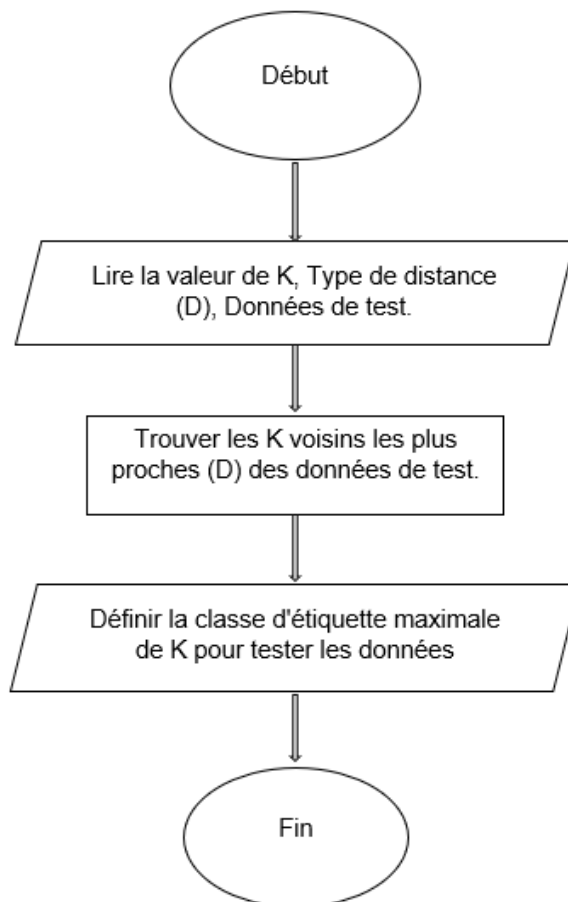


Figure 13 : Organigramme de l'algorithme KNN.

### 3.6.4. NAÏVE BAYES

Si les données ne sont pas complexes et que la tâche est relativement simple, les algorithmes Naïve Bayes peuvent être utilisés. Il s'agit d'un classificateur qui présente des avantages par rapport aux algorithmes de régression logistique et de l'algorithme k-plus proche voisin, lorsqu'il travaille avec une quantité limitée de données disponibles pour former un modèle.

L'algorithme Naïve Bayes est également un bon choix lorsque les ressources CPU et mémoire sont un facteur limitant. Parce qu'il est très simple, il n'a pas tendance à surcharger les données et peut être formé très rapidement. Il fonctionne également bien avec les nouvelles données continues utilisées pour mettre à jour le classificateur. Si la taille et la variance des données augmentent et que vous avez besoin d'un modèle plus complexe, d'autres classificateurs fonctionneront probablement mieux. De plus, sa simple analyse n'est pas une bonne base pour des hypothèses complexes.

L'algorithme Naïve Bayes est souvent le premier algorithme que les scientifiques essaient d'utiliser pour analyser du texte. C'est un algorithme de classification qui applique une estimation de densité aux données. L'algorithme exploite le théorème de Bayes et suppose que les données prédites sont conditionnellement indépendantes. C'est un classificateur probabiliste, qui est basé sur le théorème de Bayes :

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} \quad (3.16)$$

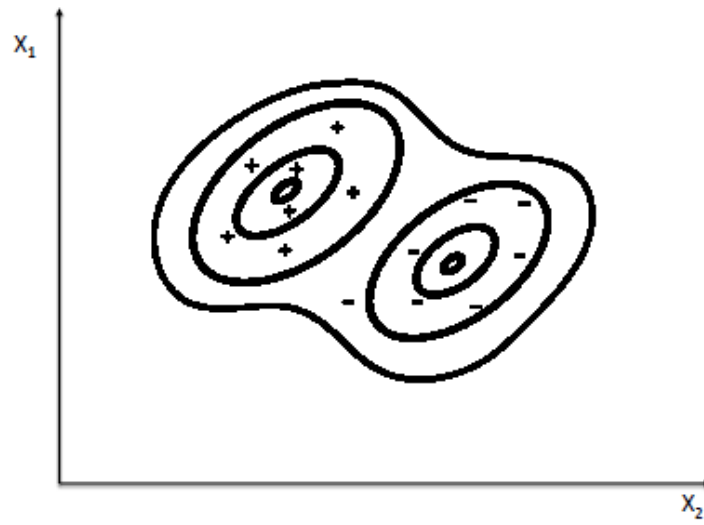


Figure 14 : Principe de l'algorithme Naïve Bayes.

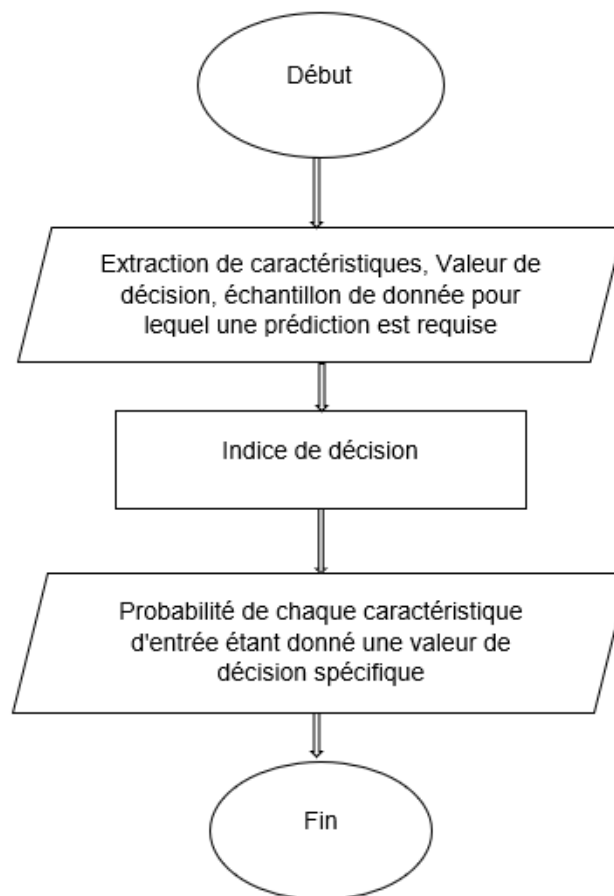


Figure 15 : Organigramme de l'algorithme Naïve Bayes.



### 3.6.5. RÉSEAUX NEURONES ARTIFICIELS (ANN)

Les réseaux neurones artificiels sont des méthodes d'apprentissage supervisé et non supervisé qui tentent d'imiter la pensée humaine par la modélisation simplifiée des systèmes neuronaux du cerveau de l'homme. L'intention n'est pas de modéliser le mécanisme exact du fonctionnement biologique des cellules neurales dans le cerveau ni de créer un clone biologique. Au lieu de cela, la biologie n'est que la source d'inspiration. L'élément de traitement unitaire est un modèle simple appelé neurone. Chaque neurone est fondamentalement une fonction pouvant recevoir plusieurs entrées et ne produire qu'une sortie. La combinaison de plusieurs neurones en un réseau de neurones est-ce que nous avons appelé réseau de neurones artificiels [50].

Ce concept fut introduit pour la première fois en 1943 par McCulloch qui a proposé la première définition de neurone formel [51]. Puis, en 1957, Rosenblatt réalisa le premier réseau avec une couche d'entrée et de sortie afin de simuler le fonctionnement rétinien dans la reconnaissance de forme [52]. Depuis cette époque et au travers des puissances computationnelles qui ne cessent d'augmenter, il a été développé un nombre non négligeable de différents types de réseaux neurones artificiels [53]. Quelques exemples d'algorithme de réseaux de neurones : McCulloch and Pitts, Perceptron, ADaptive LInear Neuron (ADALINE), Multi-Layer Perceptron (MLP), Many ADaptive LInear Neuron (MADALINE) et Back Propagation Network.

Un réseau de neurones artificiels (ANN) peut apprendre et donc être formé pour trouver des solutions, reconnaître des modèles, classer des données et prévoir des événements futurs. Les réseaux de neurones artificiels sont utilisés pour résoudre des problèmes plus complexes, tels que la reconnaissance des caractères, la prédiction boursière et la compression d'image. Le comportement d'un réseau de neurones est défini par la façon dont ses éléments individuels

sont connectés et par les forces de ces connexions, ou poids. Les poids sont automatiquement ajustés en entraînant le réseau selon une règle d'apprentissage spécifiée jusqu'à ce qu'il exécute correctement la tâche souhaitée.

Les réseaux de neurones artificiels sont excellents pour modéliser des données non linéaires avec un grand nombre de fonctionnalités d'entrée. Lorsqu'ils sont utilisés correctement, les ANN peuvent résoudre des problèmes qui sont trop difficiles à résoudre avec un algorithme simple. Cependant, les réseaux de neurones sont coûteux en calcul, il est difficile de comprendre comment un réseau de neurones artificiels a atteint une solution.

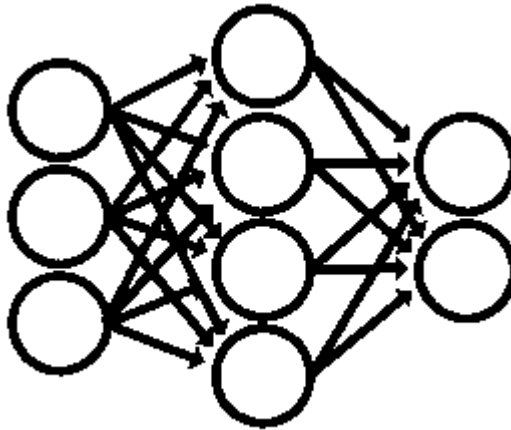


Figure 16 : Principe des réseaux de neurones artificiels.

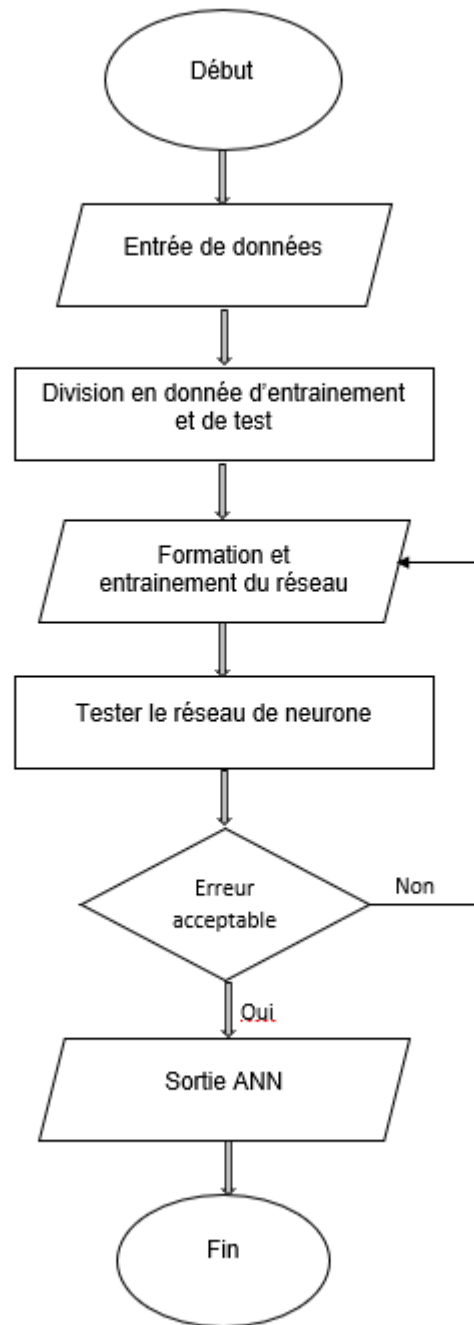


Figure 17 : Organigramme de l'algorithme ANN.

### **3.7 CONCLUSION**

À travers cette section, nous avons démontré la méthodologie utilisée pour réaliser notre projet de recherche. Il faut retenir que les principales étapes pour la classification des signaux se font par une acquisition des signaux recueillis des capteurs placés sur les équipements et suivis d'une phase de prétraitement de ces signaux recueillis. Ensuite, il faut extraire les caractéristiques qui sont indispensables pour entraîner nos algorithmes de classification. L'objectif principal dans notre travail de recherche est de se servir de cette base pour réaliser une solution qui permet d'améliorer le processus de maintenance dans les industries. Nous devons alors mettre en place une stratégie de détection des défauts qui combine une analyse des fréquences de fonctionnement grâce à notre système de surveillance de l'état de la machine et les algorithmes de classifications. Ainsi, à la section suivante nous allons présenter les différents résultats que nous avons obtenus en appliquant cette méthodologie.

## CHAPITRE 4

### ANALYSE DES RÉSULTATS ET DISCUSSION

#### 4.1. APERÇU DES SIGNAUX

En nous basant sur la méthodologie proposée au chapitre 3 et les différentes techniques d'extraction de caractéristiques connue dans la littérature, nous avons procédé par analyser notre signal de vibration recueilli du capteur qui a fourni des données défailtantes. Dans un premier temps pour réaliser notre système de surveillance de l'état de la machine nous avons extrait de ce signal la moyenne des fréquences crêtes. Étant donné que la fréquence de crête moyenne distingue avec succès les signaux nominaux des signaux défectueux, nous avons extrait la fréquence de crête moyenne de chaque segment de données de notre signal.

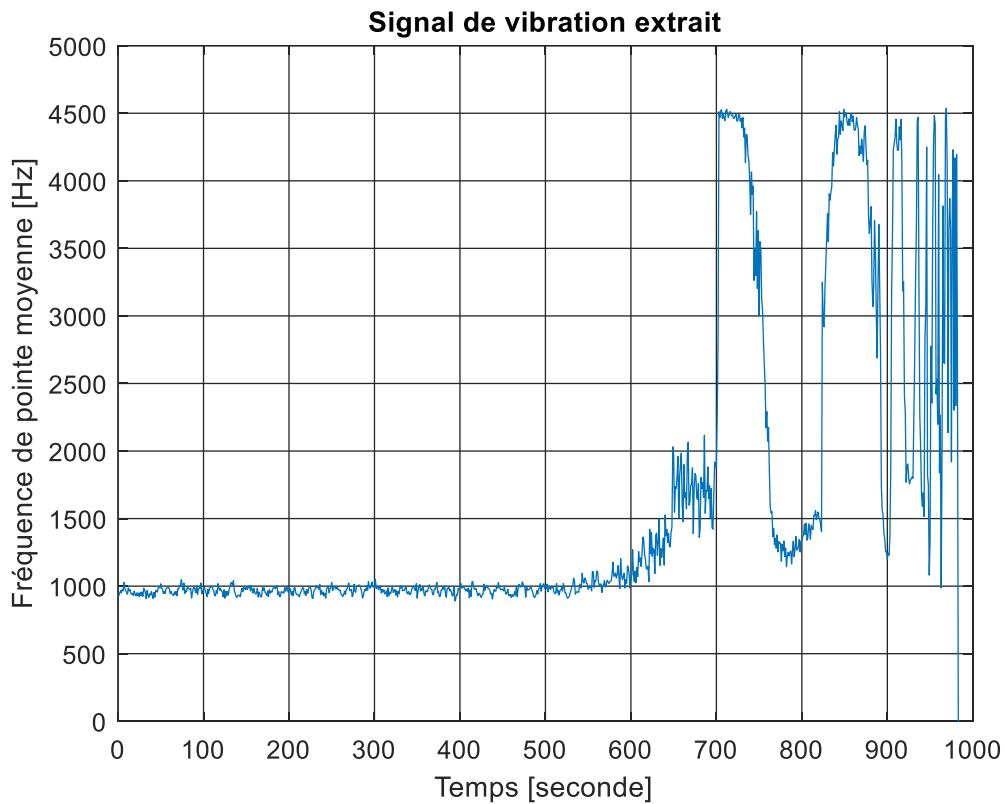


Figure 18 : Fréquences maximales moyennes extraites en fonction du temps.

## **4.2. ANALYSE DES CARACTÉRISTIQUES SÉLECTIONNÉES**

Pour réaliser la classification des signaux grâce à l'apprentissage automatique en fonction des différentes classes que nous avons définies à savoir les trois niveaux de dégradation du signal de vibration (signal nominal, signal dégradé, signal défaillant) ; il nous faut un meilleur choix dans la sélection des caractéristiques que nous allons utiliser pour entraîner nos algorithmes de classification. Pour cela après plusieurs analyses des différentes caractéristiques nous présentons dans cette section, celle que nous avons retenue (Kurtosis et RMS) pour la réalisation de notre projet de recherche dû au fait qu'elles distinguent le milieu les niveaux de dégradation par rapport aux autres caractéristiques que nous avons extraites.

À la suite de la visualisation des signaux, nous sommes passés à l'étape de sélection des caractéristiques. Tout d'abord, nous allons présenter la méthode temporelle, ensuite nous présenterons la méthode statistique.

### **4.2.1. MÉTHODE TEMPS-FREQUENCE**

Dans cette section, des caractéristiques représentatives sont extraites de chaque segment de données. Mais avant de sélectionner les fonctions à utiliser, nous allons tracer le spectrogramme des signaux de vibration. La visualisation des signaux dans le domaine temporel et fréquentiel peut aider à découvrir des modèles de signal qui indiquent une dégradation ou une défaillance.

Les représentations temps-fréquence décrivent comment le contenu spectral d'un signal évolue en fonction du temps. Ils peuvent être utilisés pour former les algorithmes qui peuvent identifier et extraire des modèles à partir des représentations temps-fréquence et également choisir parmi une variété de techniques qui peuvent générer des représentations temps-fréquence pour les signaux, notamment le spectrogramme.

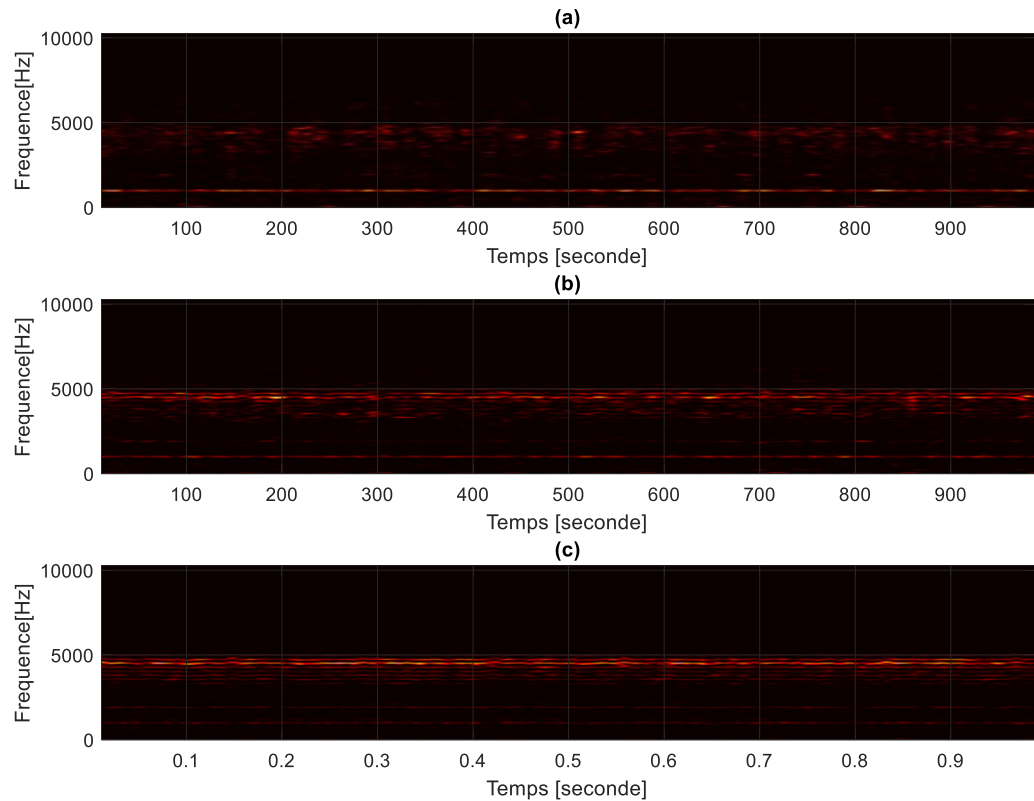


Figure 19 : Tracer du spectrogramme des signaux de vibration. (a)Signal nominal ; (b)Signal dégradé ; (c)Signal défaillant.

Avec le spectrogramme nous constatons que les trois niveaux de dégradation sont bien distingués sur la figure 15. Pour le signal nominal nous observons une concentration de la puissance du signal dans les fréquences situées au tour de 1000Hz et pour les autres types de signaux une concentration du niveau de signal dans des fréquences beaucoup plus élevées notamment pour le signal défaillant ou la majeure partie de la puissance du signal se concentre à des fréquences situées autour de 4000Hz et 5000Hz. Cela démontre l'utilité d'utiliser les méthodes temps-fréquence pour extraire les caractéristiques des signaux, car elle offre de meilleurs résultats pour entraîner nos algorithmes de classification.

#### 4.2.2. MÉTHODE STATISTIQUE

Avec la boîte à outils Diagnostic Feature Designer du logiciel MATLAB, nous avons importé et visualisé l'ensemble des données de vibration dans le but d'extraire les caractéristiques essentielles pour réaliser la classification de nos signaux avec les algorithmes d'apprentissage automatique supervisée. Nous avons obtenu les résultats représentés sur la figure 16 qui montre que les caractéristiques obtenues avec le RMS sont meilleures, car elle distingue mieux les trois groupes de données en fonction du niveau de défaillance par rapport aux caractéristiques de Kurtosis.

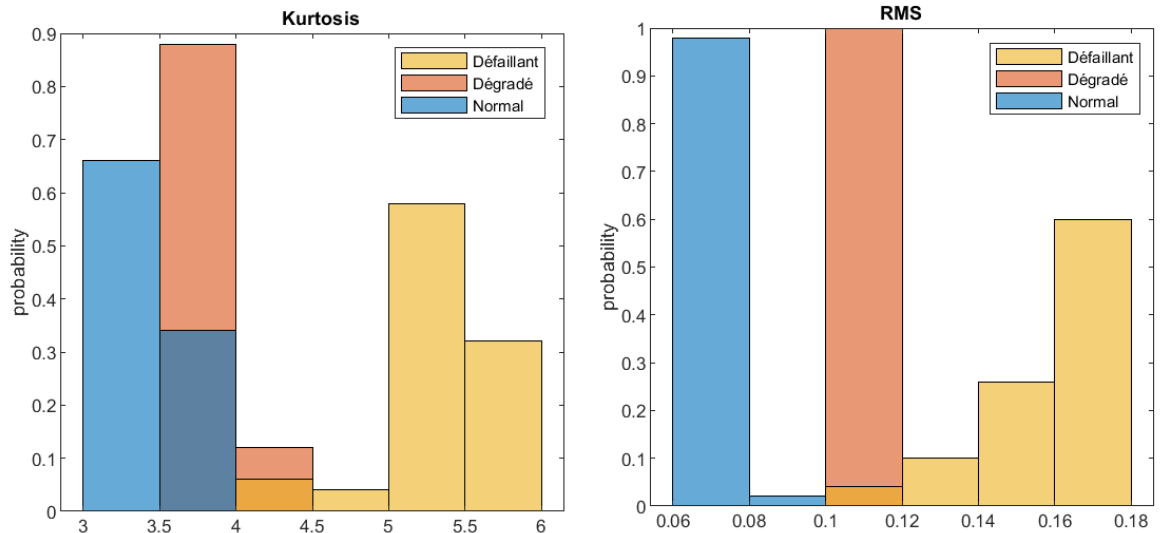


Figure 20 : Histogramme des caractéristiques des signaux de vibration.

La figure 16 nous présente l'histogramme des caractéristiques des signaux obtenues grâce à la méthode de Kurtosis et de RMS. Elle nous permet par la même occasion de comparer les deux méthodes et de constater que RMS dans le cas de notre projet offre une meilleure distinction des différents niveaux de dégradation de notre signal. Avec Kurtosis nous constatons que la probabilité pour qu'un signal nominal se situe sur une plage de valeur comprise entre [3 et 3.5] est de 65% et avec RMS cette probabilité est de 99% pour des valeurs comprises entre [0.06 et 0.08].



### 4.3. RÉSULTATS DES TESTS

Les résultats de notre projet de recherche montrent comment extraire des caractéristiques des données mesurées pour effectuer une surveillance de l'état d'une machine avec des capteurs installés pour recueillir les données de vibration. Sur la base des fonctionnalités extraites, des modèles dynamiques sont générés, validés et utilisés pour prévoir les pannes afin que des actions puissent être prises avant que les pannes réelles ne se produisent.

Dans l'objectif de mettre en place un système de prédiction de défaut des machines tournante capable de prédire les pannes avant que celle-ci ne produise et dans l'optique de fournir un système robuste et fiable, nous avons décidé de développer ce projet avec deux méthodes d'analyse de la machine. La première méthode est intuitive, car elle est basée sur une interprétation graphique de l'évolution la moyenne des fréquences crête dans le temps. Pour cela nous avons mis en place un système appelé système de surveillance de l'état de la machine qui va au fur et à mesure que la machine est en marche recevoir des données recueillies par les capteurs et mettre à jour le graphique qui représente les résultats du calcul de la moyenne des fréquences crête. Grâce aux seuils de bon fonctionnement qui sont fixés de base sur le graphique les opérateurs pourront vérifier s'il y a une augmentation des valeurs de la moyenne et prendre des décisions d'intervenir avant que la panne ne survienne.

Cette première méthode a été développée dans le but de soutenir la seconde, qui est l'application de l'apprentissage automatique sur les signaux de nos capteurs plus précisément les algorithmes de classification des signaux détaillés dans la section de la méthodologie. Avec la combinaison de ces deux méthodes, nous serons plus aptes à interpréter et à valider un signal de dégradation ou de panne de la machine comme une vraie défaillance en limitant au maximum les risques de se tromper.

#### 4.3.1. RÉSULTATS DU SYSTÈME DE SURVEILLANCE DE L'ÉTAT DE LA MACHINE

Dans cette section, la surveillance de l'état de la machine est effectuée à l'aide des seuils prédéfinis et des modèles dynamiques. Pour surveiller les conditions de fonctionnement de la machine, nous avons créé une alarme qui se déclenche si la fréquence de crête moyenne dépasse le seuil prédéfini. La prévision nous aide à mieux nous préparer à une panne potentielle ou même à arrêter la machine avant une panne. La figure 17 montre que les valeurs prévues de la fréquence de crête moyenne sont bien inférieures au seuil et au fur et à mesure que de nouvelles données arrivent, le système met à jour les paramètres du modèle et estime les valeurs en fonction des seuils prédéfinis. Une fois, le seuil atteint une alarme est automatiquement générée pour arrêter la machine ou informer les opérateurs d'un potentiel défaut.

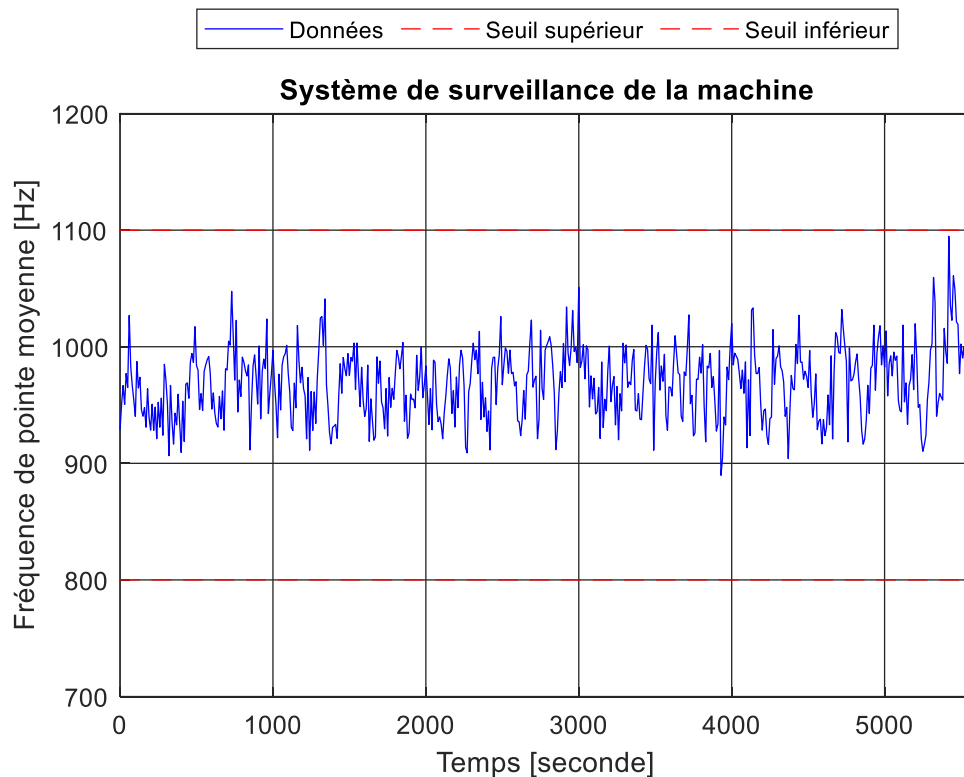


Figure 21 : Mise en œuvre du système de surveillance avec détection de panne.

#### 4.3.2. RÉSULTATS DES TESTS DE CLASSIFICATION DES SIGNAUX

Une étape importante lors de l'utilisation de l'apprentissage automatique consiste à vérifier les performances des modèles que nous avons utilisés. L'évaluation d'un système de classification se base généralement sur un partage de la base de données en trois ensembles aussi appelés sous-ensemble :

- Un sous-ensemble de donnée d'apprentissage pour la conception du système de classification, c'est un jeu de données d'exemples utilisés pour l'apprentissage.
- Un sous-ensemble de données de test qui est indépendant de l'ensemble de données de formation, mais qui suit la même distribution de probabilité pour évaluer l'architecture du système de classification.
- Un sous-ensemble de données de validation qui est un ensemble de données exemples utilisés pour ajuster les hyperparamètres du classificateur et pour valider la pertinence de l'architecture choisie.

La distinction entre les deux premiers ensembles est nécessaire pour évaluer les performances du système de classification en ne testant pas le système avec des échantillons ayant servi pour sa conception. Le troisième sous-ensemble permet d'obtenir une validation indépendante des critères de sélection. Dans certains cas, la taille de la base de données ne permet pas d'avoir un partage en trois sous-ensembles ; dans ce cas on utilise uniquement un partage en deux : un sous-ensemble d'apprentissage et un sous-ensemble de tests. Il existe différents types de méthode pour réaliser le partage de la base de données en ensemble d'apprentissages et en ensemble de tests. Parmi ces méthodes, on trouve la validation croisée (k-fold) et la validation par holdout.

#### **4.3.2.1. MÉTHODE DE VALIDATION CROISÉE (K-FOLD)**

Une méthode d'évaluation des performances d'un algorithme d'apprentissage automatique est la validation croisée. Cette technique permet à l'algorithme de faire des prédictions en utilisant des données non utilisées pendant la phase de formation. La validation croisée partitionne un ensemble de données et utilise un sous-ensemble pour entraîner l'algorithme et les données restantes à tester. Étant donné que la validation croisée n'utilise pas toutes les données pour créer un modèle, il s'agit d'une méthode couramment utilisée pour éviter le surapprentissage pendant la formation.

Le principe de cette méthode de validation consiste à partitionner les données en  $k$  sous-ensembles (ou plis) choisis au hasard de taille à peu près égale. Un sous-ensemble est utilisé pour valider le modèle formé à l'aide des sous-ensembles restants. Ce processus est répété  $k$  fois, de sorte que chaque sous-ensemble soit utilisé exactement une fois pour la validation. L'ensemble de formation est ensuite utilisé pour former un algorithme d'apprentissage supervisé, et l'ensemble de tests est utilisé pour évaluer ses performances. Ce processus est répété plusieurs fois et l'erreur de validation croisée moyenne est utilisée comme indicateur de performance.

#### **➤ RÉSULTATS DES TESTS AVEC LA VALIDATION CROISÉE**

Avec cette approche, nous avons obtenu un taux de classification de 99.3% avec l'algorithme de Naïve Bayes qui est le meilleur parmi l'ensemble des algorithmes testés. La figure 18 nous montre les groupes de classification créée avec ce classificateur.

Pour réaliser les tests de validation avec la méthode de validation croisée, nous avons choisi de configurer nos paramètres de validation de la façon suivante : pour un ensemble de (403\*20480 lignes) données, nous avons utilisé un paramètre de K égal à 5 pour avoir les meilleures performances de nos algorithmes de classification.

Avec ces paramètres, l'ensemble de données a été divisé en (403 diviser par 5) sous-ensemble pour servir de données d'apprentissage et de donnée de test. Et le résultat nous permet de constater sur la figure 18 un groupement en trois classes de nos données toujours en respectant les valeurs de Kurtosis et de RMS que nous avons obtenues lors de l'extraction de nos caractéristiques. Et avec le taux de réussite de 99.3%, nous pouvons conclure que le choix des caractéristiques utilisées est le meilleur choix. Mais toujours dans l'optique d'améliorée se taux de réussite nous avons opté de tester la deuxième méthode de validation Holdout avec des paramètres qui nous fourniront un meilleur taux de réussite.

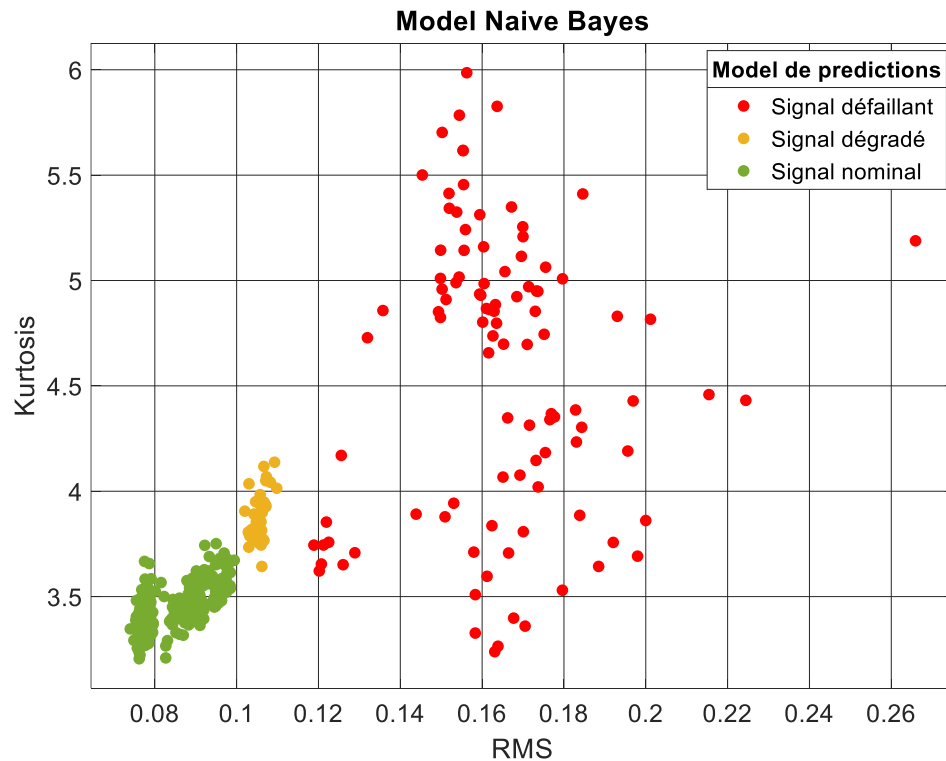


Figure 22 : Classification des signaux avec la méthode de validation croisée.

#### **4.3.2.2. MÉTHODE DE VALIDATION HOLDOUT**

Le holdout, c'est quand vous divisez votre ensemble de données en un ensemble « train » et « test ». L'ensemble d'apprentissages est ceux sur quoi le modèle est formé, et l'ensemble de tests est utilisé pour voir dans quelle mesure ce modèle fonctionne sur les données. Une division commune lors de l'utilisation de la méthode de holdout consiste à utiliser 80% des données pour la formation et les 20% restants des données pour les tests. Le principe de cette méthode de validation consiste à partitionner les données en exactement deux sous-ensembles d'un rapport spécifié pour la formation et la validation.

#### **➤ RÉSULTATS DES TESTS AVEC LA VALIDATION HOLDOUT**

La performance globale de cette approche proposée est la meilleure. Dans cette approche, nous avons obtenu un taux de classification de 100% avec tous les algorithmes testés. La figure 19 nous montre les groupes de classification créée avec cette méthode.

Pour réaliser les tests de validation avec la méthode de validation holdout, nous avons choisi de configurer nos paramètres de validation de la façon suivante : Pour un ensemble de (403\*20480 lignes) données nous avons partitionné les données en deux groupes avec un pourcentage de 70% pour les données d'entraînement et un pourcentage de 30% qui seront utiliser comme données de test pour nos algorithmes de classification.

Avec ces paramètres l'ensemble de données a été divisé en (70% de 403 égales à 283) sous-ensemble pour servir de données d'apprentissage et en (30% de 403 égales à 120) de donnée de test. Et le résultat nous permet de constater sur la figure 19 un groupement en trois classes de nos données toujours en respectant les valeurs de Kurtosis et de RMS que nous avons obtenues lors de l'extraction de nos caractéristiques. Et avec un taux de réussite cette fois de 100%, nous pouvons conclure que le choix des caractéristiques utiliser est le meilleur

choix et que la validation Holdout avec des paramètres que nous avons utilisés offre un meilleur taux de réussite.

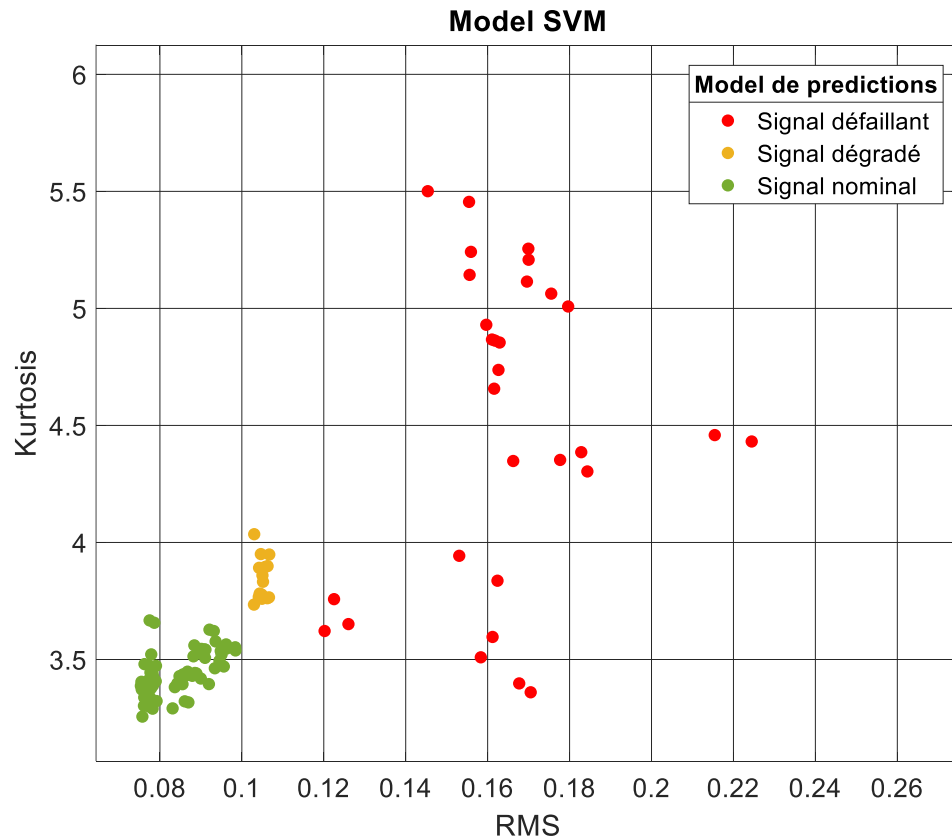


Figure 23 : Classification des signaux avec la méthode de la validation holdout.

#### 4.3.2.3. MATRICE DE CONFUSION

L'évaluation d'un système de classification se base généralement sur la mesure du taux de classification. Pour avoir une meilleure interprétation des résultats, nous utilisons la matrice de confusion. Il s'agit d'une matrice carrée de taille  $[M \times M]$  où  $M$  représente le nombre de classes. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe prédite, alors que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence).

La matrice de confusion offre la possibilité d'analyse du taux de bonne classification (ou d'erreur) par classe. Les valeurs de la diagonale de la matrice représentent le nombre de formes correctement classifiées. La somme des valeurs par ligne correspond au nombre d'échantillons de test par classe. Le taux de classification par classe est donné par la valeur à la diagonale divisée par la somme des valeurs par ligne.

➤ **MATRICE DE CONFUSION AVEC LA VALIDATION CROISÉE**

**Model Naive Bayes**

Vraie classe	Signal défaillant	101		
	Signal dégradé		51	
	Signal nominal		3	248
		Signal défaillant	Signal dégradé	Signal nominal
		Classe Prédicte		

Figure 24 : Matrice de confusion avec la méthode de validation croisée.

La figure 20 nous présente la matrice de confusion obtenue avec la méthode de validation croisée. Nous pouvons valider le résultat du taux de réussite de 99.3% obtenue avec cette méthode, car avec un nombre d'échantillons de 251 du signal nominal utilisé, nous avons obtenu une reconnaissance de 248 échantillons et 3 échantillons du signal nominal ont été confondus comme échantillons de signal dégradé. Pour le reste, des échantillons ils ont très bien été repartis par l'algorithme de classification Naïve Bayes utiliser avec la validation croisée.



➤ **MATRICE DE CONFUSION AVEC LA VALIDATION HOLDOUT**

		Model SVM		
Vraie classe	Signal défaillant	30		
	Signal dégradé		15	
	Signal nominal			75
		Signal défaillant	Signal dégradé	Signal nominal
		Classe prédicte		

Figure 25 : Matrice de confusion avec la méthode de la validation holdout.

La figure 21 nous présente la matrice de confusion obtenue avec la méthode de validation holdout. Nous pouvons valider le résultat du taux de réussite de 100% obtenue avec cette méthode, car avec un nombre de 75 échantillons du signal nominal utilisé nous avons une reconnaissance des 75 échantillons comme échantillons de signal nominal. Et ces pareils avec les deux autres classes de signaux, nous observons une réussite de classification de 100% aussi. Donc l'ensemble des échantillons ont très bien été répartis par l'algorithme de classification SVM utiliser avec la validation holdout.

La méthode de validation holdout est généralement utilisée lorsqu'on dispose d'un très grand ensemble de données pour construire le modèle initial du projet de classification de signaux. Et la validation croisée utilise plusieurs fractionnements d'entraînement et de test, il faut donc plus de puissance de calcul et de temps pour l'exécution lors de l'utilisation de la méthode de validation croisée.

Les tableaux 4 et 5 ci-dessous montrent la comparaison des taux de réussites entre les différents types d'algorithmes de classification que nous avons entraînés et testés. Nous pouvons donc constater qu'avec la méthode de la validation croisée sur le tableau 4, nous avons obtenu un taux de 98.7% pour l'algorithme arbre de décision et un taux de 99.3% pour les autres types d'algorithmes à savoir le SVM, KNN et Naïve Bayes. Et pour la méthode de validation holdout sur le tableau 5, nous avons obtenu un taux de 100% pour les différents types d'algorithmes utilisés.

**Tableau 4 : Résultats des tests avec la validation croisée**

Algorithmes	Résultats
Arbre de décision	98.7%
Naïve Bayes	99.3%
SVM	99.3%
KNN	99.3%

**Tableau 5 : Résultats des tests avec la validation holdout**

Algorithmes	Résultats
Arbre de décision	100%
Naïve Bayes	100%
SVM	100%
KNN	100%

## CHAPITRE 5

### CONCLUSION ET PERSPECTIVES

#### 5.1. CONCLUSION

Le présent travail a porté sur la conception d'un outil de surveillance de l'état d'une machine tournante et d'un outil de classification de signaux grâce à l'apprentissage automatique. Une étude de la littérature a été menée dans un premier temps. Et cela nous a permis d'identifier les technologies et outils existants pour surveiller les machines et les algorithmes d'apprentissage machine disponible pour analyser les signaux vibrations.

Le projet de recherche présenté dans cette étude propose, une solution permettant d'exploiter le potentiel des modèles d'intelligence artificielle pour l'amélioration du processus de maintenance dans l'industrie par la reconnaissance des défauts des machines. Le contexte décrit dans le chapitre 1 montre l'importance de posséder un programme de maintenance prédictive efficace. En effet, les arrêts non planifiés des chaînes de production pour cause de bris sont des sources de perte financière et de perte de temps affectant directement la santé budgétaire et la compétitivité des entreprises.

En conclusion, les informations qui sont obtenues grâce aux systèmes de surveillance et de l'apprentissage automatiques peuvent aider les responsables de la maintenance à préparer les ressources matérielles et humaines nécessaires avant que survienne une défaillance. Ainsi, les stratégies de maintenance traditionnelles impliquant une maintenance corrective et préventive peuvent être remplacées par une maintenance prédictive. Avec un excellent taux de reconnaissances à 100 %, cela démontre que l'utilisation de la reconnaissance des signaux de vibration peut permettre une prédiction plus précise des différents éléments en défaillances sur les machines. Par la suite de ce chapitre, nous allons mettre en lumière les perspectives d'amélioration et les apports d'un point de vue personnel de notre projet de recherche.

## **5.2. PERSPECTIVE**

Malgré la qualité des recherches et le succès de l'approche proposée dans ce mémoire, il subsiste plusieurs limitations. Toutefois, pour améliorer notre système qui présente une insuffisance due au fait que nous avons travaillé sur des données expérimentales et non sur des données obtenues sur des machines en pleine production, il faudra procéder par une phase pilote de test en situation réelle avant de se lancer dans l'implémentation du système dans un milieu industriel.

De plus, durant ce projet de recherche, nous avons acquis de nouvelles connaissances en intelligence artificielle avec les algorithmes de classifications. Cette expérience nous a également donné la possibilité d'améliorer nos compétences en méthodologie de recherche, en communication et en rédaction. Nous avons aussi travaillé sur un article scientifique qui a été accepté et sera présenté lors d'une conférence internationale traitant des mêmes thématiques et sera publié dans un journal pour évaluation de notre projet de recherche.

Pour finir, cette maîtrise nous a donné l'envie de poursuivre dans le domaine de la recherche et de l'enseignement universitaire avec pour objectif de contribuer aux avancées scientifiques dans le monde et d'apprendre toujours plus.

## BIBLIOGRAPHIE OU LISTE DE RÉFÉRENCES

- [1] R. CGI, "Industrie 4.0 : Pour une entreprise hautement concurrentielle," *CGI group inc*, 2017.
- [2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, et M. Gidlund, "Industrial internet of things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724-4734, 2018.
- [3] D. Aschenbrenner, M. E. Latoschik, et K. Schilling, "Industrial maintenance with augmented reality: two case studies," Dans *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 341-342, 2016.
- [4] T. ING, "Surveillance vibratoire et maintenance prédictive Techniques de l'Ingénieur," *Techniques de l'ingénieur Vibrations en milieu industriel*, vol. 6, pp. 100-2005.
- [5] M. Jacques, "Surveillance vibratoire et maintenance prédictive," (en fre), *Techniques de l'ingénieur Vibrations en milieu industriel, mesures, surveillance et contrôle*, Article de base documentaire vol. base documentaire : TIB424DUO, no. ref. article : r6100, 2002.
- [6] D. Klitou, Conrads, J. and Rasmussen, M., "Germany: Industrie 4.0. 1st ed. [ebook] European Commission, pp.2-7.," *DIGITAL TRANSFORMATION MONITOR*, 2017.
- [7] W. Ashraf, "A model to assess and customize computerized maintenance management systems complies with industry 4.0 vision and requirements: A case study in food processing plant.," *Master's thesis in Offshore technology : Industrial asset management*, p. 102, 2018.
- [8] D. S. K. e. C. Rostetter, "MAINTENANT DIGITALE PRESCRIPTIVE Internet des objets, Big Data et Dynamic Case Management: un véritable tsunami au niveau des flux de valeur dans le domaine de la fabrication," *Pegasystems Inc., États-Unis*, p. 20.
- [9] H. Q. J. Lee, G. Yu, J. Lin. *Bearing Data Set (Rexnord Technical Services éd.)*, 2007.
- [10] U. i. d. t. UIT, "SÉRIE Y: INFRASTRUCTURE MONDIALE DE L'INFORMATION, PROTOCOLE INTERNET ET RÉSEAUX DE PROCHAINE GÉNÉRATION. Réseaux de prochaine génération – Cadre général et modèles architecturaux fonctionnels. Présentation générale de l'Internet des objets " *SECTEUR DE LA NORMALISATION DES TÉLÉCOMMUNICATIONS DE L'UIT*, p. 24, 2012.
- [11] D. Sembroiz, S. Ricciardi, et D. Careglio, "Chapter 10 - A Novel Cloud-Based IoT Architecture for Smart Building Automation," Dans *Security and Resilience in Intelligent Data-Centric Systems and Communication Networks*, M. Ficco and F. Palmieri, Éd.: Academic Press, pp. 215-233, 2018.
- [12] P. P. Ray, "A survey on Internet of Things architectures," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 3, pp. 291-319, 2018.
- [13] N. Rakesh, B. U. Maheswari, et S. K. Srivatsa, "Performance analysis of propagation models in different terrain conditions for IEEE standard 802.16e WiMAX," Dans *2014 International Conference on Communication and Signal Processing*, pp. 142-146, 2014.
- [14] A. Kumar et G. P. Hancke, "Energy efficient environment monitoring system based on the IEEE 802.15. 4 standard for low cost requirements," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2557-2566, 2014.
- [15] I. Lee et K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431-440, 2015.

- [16] Hewlett-Packard, "Une étude de HP révèle que 70% des périphériques de l'Internet des objets sont vulnérables aux attaques

Les appareils IoT ont présenté en moyenne 25 vulnérabilités par produit, ce qui indique une surface d'attaque étendue pour les adversaires," 2014.

- [17] L. C. De Silva, C. Morikawa, et I. M. Petra, "State of the art of smart homes," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1313-1321, 2012.
- [18] D. Miorandi, S. Sicari, F. De Pellegrini, et I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497-1516, 2012.
- [19] S. Kitagami, V. T. Thanh, D. H. Bac, Y. Urano, Y. Miyanishi, et N. Shiratori, "Proposal of a Distributed Cooperative IoT System for Flood Disaster Prevention and its Field Trial Evaluation," *International Journal of Internet of Things*, vol. 5, no. 1, pp. 9-16, 2016.
- [20] A. Alphonsa et G. Ravi, "Earthquake early warning system by IOT using Wireless sensor networks," Dans *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1201-1205: IEEE, 2016.
- [21] N. Fatema et R. Brad, "Attacks and counterattacks on wireless sensor networks," *arXiv preprint arXiv:1401.4443*, 2014.
- [22] G. Cerullo, G. Mazzeo, G. Papale, B. Ragucci, et L. Sgaglione, "Chapter 4 - IoT and Sensor Networks Security," Dans *Security and Resilience in Intelligent Data-Centric Systems and Communication Networks*, M. Ficco and F. Palmieri, Éd.s.: Academic Press, pp. 77-101, 2018.
- [23] S. Sakr, A. Liu, D. M. Batista, et M. Alomari, "A survey of large scale data management approaches in cloud environments," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 311-336, 2011.
- [24] L. Wang et X. V. Wang, *Cloud-Based Cyber-Physical Systems in Manufacturing* (Springer). Springer: Springer, 2018.
- [25] A. T. Sylvie Rioux et Bertrand Gauvreau, "Comment le numérique transformera la gestion de votre usine ?," *Andromedia Technologies*, 2019.
- [26] A. Villemeur, *Sûreté de fonctionnement des systèmes industriels : Fiabilité-facteurs humains informatisation*. Paris: Ed. Eyrolles, 1988.
- [27] T. Stéphane, "DATA MINING ET STATISTIQUE DÉCISIONNELLE La science des données," *Editions TECHNIP*, vol. 5e edition, p. 934, 2017.
- [28] T. Abdessettar, "DÉVELOPPEMENT D'UN SYSTÈME DE CLASSIFICATION AUTOMATIQUE DE L'ACTIVITÉ PHYSIQUE," TELUQ, 2018.
- [29] MathWorks, "Decision Models for Fault Detection and Diagnosis," *MathWorks*.
- [30] O. D. Lara et M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192-1209, 2013.
- [31] Y. L. He, Ye, *Physical Activity Recognition Utilizing the Built-In Kinematic Sensors of a Smartphone* (International Journal of Distributed Sensor Networks). 2013.
- [32] A. A.-A.-U. K. Tonello, Institute of et al., "Machine Learning Tips and Tricks for Power Line Communications," *IEEE Access*, p. 18, 2019.
- [33] E. Alpaydin, *Introduction to machine learning* (MIT press). MIT press, 2009.
- [34] J. Yu, "Health Condition Monitoring of Machines Based on Hidden Markov Model and Contribution Analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2200-2211, 2012.

- [35] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, et G. Tripot, "A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models," *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 491-503, 2012.
- [36] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, et G. Tripot, *Hidden Markov Models for failure diagnostic and prognostic*, pp. 1-8, 2011.
- [37] F. Wang, Y. Zhang, B. Zhang, et W. Su, "Application of Wavelet Packet Sample Entropy in the Forecast of Rolling Element Bearing Fault Trend," présenté à Proceedings of the 2011 International Conference on Multimedia and Signal Processing - Volume 02, pp. 12-16, 2011.
- [38] M. Mortada et S. Yacout, "cbmLAD - using Logical Analysis of Data in Condition Based Maintenance," Dans *2011 3rd International Conference on Computer Research and Development*, vol. 4, pp. 30-34, 2011.
- [39] MathWorks. (2019). *Signal Processing Toolbox*.
- [40] MathWorks. (2019). *Condition Monitoring and Prognostics Using Vibration Signals*.
- [41] J. Mathew, "Common vibration monitoring techniques," Dans *Handbook of condition monitoring*: Springer, pp. 303-323, 1998.
- [42] N. Tandon et A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology international*, vol. 32, no. 8, pp. 469-480, 1999.
- [43] S. Sassi, B. Badri, et M. Thomas, "TALAF and THIKAT as innovative time domain indicators for tracking BALL bearings," Dans *Proceedings of the 24nd Seminar on machinery vibration, Canadian Machinery Vibration Association, éditeur M. Thomas, Montréal, Canada*, pp. 404-419, 2006.
- [44] W. A. Smith et R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100-131, 2015.
- [45] O. Seryasat, M. A. Shoorehdeli, M. Ghane, J. Haddadnia, et M. Zeinali, "Intelligent fault detection of ball bearing using FFT, STFT energy entropy and RMS," *Life Science Journal*, vol. 9, no. 3, pp. 1781-1786, 2012.
- [46] W. H. L. K. S. Hao et C. Peng, "Feature extraction method based on Pseudo-Wigner-Ville distribution for rotational machinery in variable operating conditions," *Chinese Journal of Mechanical Engineering*, vol. 24, no. 4, p. 1, 2011.
- [47] P. K. Kankar, S. C. Sharma, et S. P. Harsha, "Fault diagnosis of ball bearings using continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300-2312, 2011.
- [48] S. Fu, K. Liu, Y. Xu, et Y. Liu, "Rolling Bearing Diagnosing Method Based on Time Domain Analysis and Adaptive FuzzyC-Means Clustering," vol. 2016, pp. 1-8, 2016.
- [49] B. E. Boser, I. M. Guyon, et V. N. Vapnik, "A training algorithm for optimal margin classifiers," Dans *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144-152.
- [50] P. Kardi Teknomo, "K-Mean Clustering algorithm."
- [51] W. S. McCulloch et W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115-133, 1943.
- [52] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [53] A. K. Jain, J. Mao, et K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31-44, 1996.

## ANNEXES

### ANNEXE 1 : ARTICLE SCIENTIFIQUE

Article de conférence publié à la 12th International Conference on Intelligent Decision Technologies (IDT-20), Croatie, 2020.

## **Automatic Classification of Rotating Machinery Defects using Machine Learning (ML) Algorithms**

Wend-Benedo Zoungrana<sup>1</sup>, Abdellah Chehri<sup>1</sup>, Alfred Zimmermann<sup>2</sup>

<sup>1</sup> Department of Applied Sciences, University of Québec in Chicoutimi,  
Chicoutimi, QC, Canada, G7H 2B1  
wend-benedo.zoungranal@uqac.ca ; achehri@uqac.ca

<sup>2</sup> Reutlingen University, Germany Reutlingen University, Faculty of Informatics,  
Alteburgstraße 150, 72762 Reutlingen, Germany  
alfred.zimmermann@reutlingen-university.de

**Abstract.** Electric machines and motors have been the subject of enormous development. New concepts in design and control allow expanding their applications in different fields. The vast amount of data has been collecting almost in any domain of interest. They can be static; that is to say, they represent real-world processes at a fixed point of time. Vibration analysis and vibration monitoring, including how to detect and monitor anomalies in vibration data are widely used techniques for predictive maintenance in high-speed rotating machines. However, accurately identifying the presence of a bearing fault can be challenging in practice, especially when the failure is still at its incipient stage, and the signal-to-noise ratio of the monitored signal is small. The main objective of this work is to design a system that will analyze the vibration signals of a rotating machine, based on recorded data from sensors, in the time/frequency domain. As a consequence of such substantial interest, there has been a dramatic increase of interest in applying machine learning (ML) algorithms to this task. An ML system will be used to classify and detect abnormal behavior and recognize the different levels of machine operation modes (normal, degraded, and faulty). The proposed solution can be deployed as predictive maintenance for Industry 4.0.



**Keywords.** Classification, Machine Learning (ML) Algorithms, SVM, K-Nearest Neighbour, Decision Trees, Artificial Intelligence, Naive Bayes, Predictive Maintenance, Industry 4.0.

## 1 Introduction

Rotating machines represent the most significant part of the mechanisms created by designers. The rotational movement can be used to store energy, for example, in a flywheel, to transfer power, by activating a belt or a gearbox, or to recover kinetic energy from external sources, for example, thermal energy using a turbine, or the wind using a wind turbine. The rotating parts, also known as "rotors," obviously play a central role in the processes mentioned and represent the primary source of nuisance in these systems. However, several geometric or material faults, or even the interfaces, cause a loss of power in various forms. As in all mechanisms, thermal dissipation is observed, but the mechanical energy losses are also often significant, and a certain number of vibrations accompanies the movement of the rotor. First, the vibrations of the rotor can be excited, in the axial, radial directions, or torsion to its axis of rotation. Besides, oscillations also appear at the level of the guide supports (so-called bearings). Through these biases, energy waves are therefore transmitted to the non-rotating parts, or even to neighboring equipment. They are considered important components in many industrial applications as power systems, manufactories, power plants, electric vehicles, and home appliances. For certain applications, these machines may operate under unfavorable conditions, such as high temperature, dust (mining applications), which can eventually result in motor malfunctions that lead to high maintenance costs and safety hazards [1]–[3].

The integration of different sensors, the Internet of things (IoT), and intelligent software into electric machines facilitate diagnostics and report technical problems in advance. The integration of intelligent software improves the proactivity of maintenance [4]–[7]. Predictive analyzes identify the potential options to be favored for anticipation purposes. The predictive maintenance philosophy consists of scheduling maintenance activities only when a functional failure is detected [4]. The machine would then be shut down at a time when it is most convenient, and the damaged components would be replaced. If left unattended, these failures could result in costly secondary failures. One of the advantages of this approach is that the maintenance events can be scheduled in an orderly fashion. With predictive maintenance, devices participate proactively in their own maintenance in collaboration with operators.

Nowadays, significant progress in vibration signal analysis techniques used for rotating machinery monitoring [8]–[11]. Vibration analysis is one of the most widely used techniques for predictive maintenance in high-speed rotating machines. Since the bearing is the most vulnerable component in a motor drive system, accurate bearing fault diagnostics has been a research frontier for engineers and scientists for the past decades. However, accurately identifying the presence of a bearing fault can be challenging in practice, especially when the fault is still at its incipient stage, and the signal-to-noise ratio of the monitored signal is small.

In fact, there may exist many unique features or patterns hidden in the data themselves that can potentially reveal a bearing fault. Consequently, almost impossible for humans to identify these convoluted features through manual observation or interpretation. Therefore, the monitoring

requires an intelligent method to analyze vibration data generated by the sensors. An important step involved in intelligent data analysis is to identify the signal signature and used the signatures in the learning method.

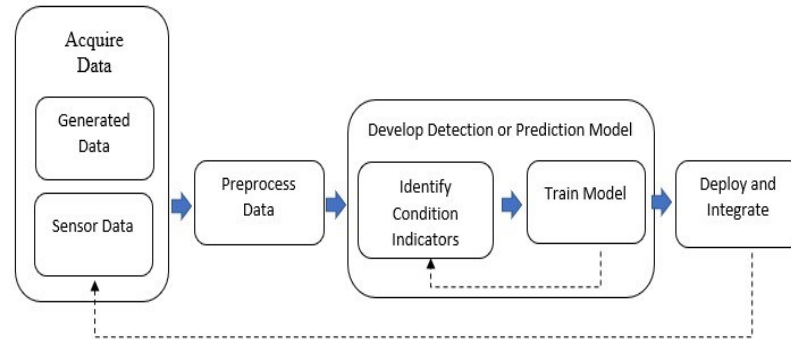
In this work, we use machine learning classification methods to recognize the different levels of machine operation modes. The goal is to recognize the normal and faulty signals from the extracted vibration signal.

The rest of this paper is organized as follows. In the next section, the motivation and related works are described. Section 3 describes the methodology. The predictive maintenance of rotating machines is described in Section 4. Section 5 provides the simulation results. Section 6 concludes this paper.

## 2 Methodology

By using sensors to determine when equipment verification is required, it is possible to prevent breakdowns and reduce routine maintenance costs. Thanks to integrated sensors connected to the Internet, control of production equipment is carried out remotely and in real-time. In this case, the recommendations are sent to the operators to remedy the problems before they even occur. This method reduces operating and capital costs by promoting repair and proactive maintenance of equipment to improve capacity utilization and productivity.

The developed method is tested on reference data extracted from the “NASA prognostic data repository” and relating to several experiences of bearing failures carried out under different operating conditions. Besides, the method is compared to traditional forecasts of time and frequency characteristics, the simulation results. Figure 1 shows the main steps of our methodology.

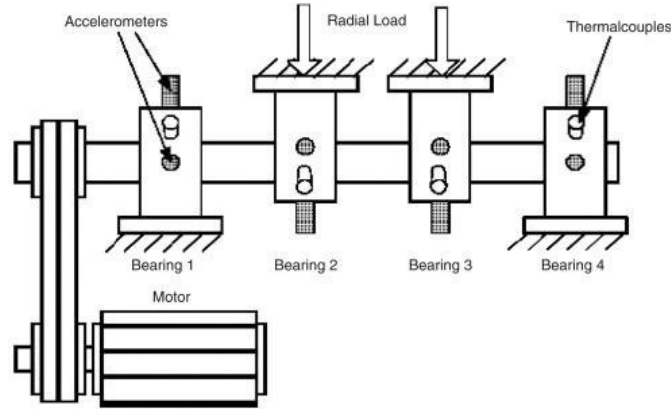


**Fig. 1.** Data analysis steps (classification).

### 2.1 Data Acquisition Protocol

Data was generated by the National Science Foundation (NSF) Industry-University Cooperative Research Centers (IUCRC) for Intelligent Maintenance Systems (IMS - [www.imscenter.net](http://www.imscenter.net)) with support from Rexnord Corporation in Milwaukee, Wisconsin, United States United. Four bearings were installed. The rotational speed was kept constant at 2000 rpm by an alternating current. The motor is coupled to the shaft via friction belts. All bearings are forcibly lubricated. Rexnord ZA-

2115 double row bearings were installed on the shaft, as shown in Figure 02. A PCB 353B33 with high sensitivity ICP Quartz accelerometers was installed on the bearing (two accelerometers for each bearing for data set 1, an accelerometer for each bearing for data sets 2 and 3). Figure 2 also shows the position of the sensors. All failures occurred after exceeding the expected lifetime of the bearing, which is more than 100 million rotations [12].



**Fig. 2.** Illustration of the bearing test ring and sensor placement- IMS dataset [12].

## 2.2 Data pre-processing

Pre-processing condition monitoring data is a very important and fundamental step when developing data models systems.

The first step in the pre-processing data stage is to filter out noise from the data. We have proceeded by applying a median filter to the vibration signals to keep the useful information and by eliminating the high-frequency noise components.

Since the spectrograms  $P(t, \omega)$  of the data for good and defective bearings are different, representative characteristics can be extracted from the spectrograms and used to accurately calculate the average peak frequency as an indicator of the state of the system.

Once the data filtering is complete, the next step is to segment the data to eliminate the nonlinear effects. There is three distinct speed region: (1) normal mode, (2) degraded mode and (3) faulty mode. Figure 3 shows the extracted signal for different modes (in the time domain).

## 2.3 Time signal based approach

Time series data in their nature is way more complicate than static data. Therefore, it is more challenging to get insightful knowledge from them. As a consequence, the ability of time series algorithms to provide comprehensible classification results becomes extremely important.

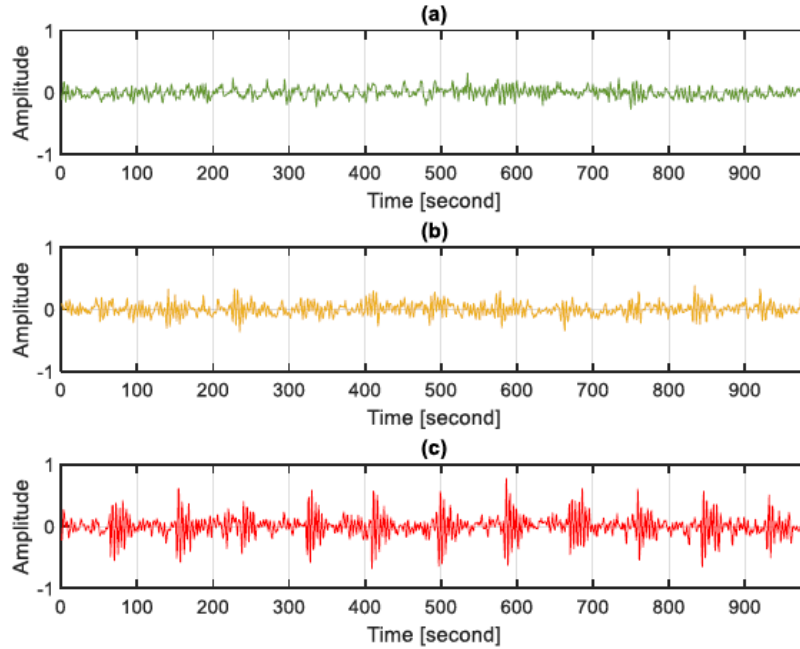
The temporal characteristics capable of identifying the precise location of the faults are as follows:

**Root Mean Square (RMS):** RMS relates to the power of the vibration signal. RMS is sensitive to load and speed changes (Eq. 1). RMS indicates the general condition at the last stages of degradation, and it is one of the important factors for machinery status diagnosis.

$$V_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_t(i)^2} \quad (1)$$

**Crest Factor (CF):** Indicates a relationship between the peak value of the signal and the RMS value indicating early signs of damage, especially when the vibration signals have impulsive characteristics (Eq. 2). It is used to determine the deterioration of bearings by relative comparison.

$$CP = \frac{\max(x_t(i))}{V_{RMS}(x_t(i))} \quad (2)$$

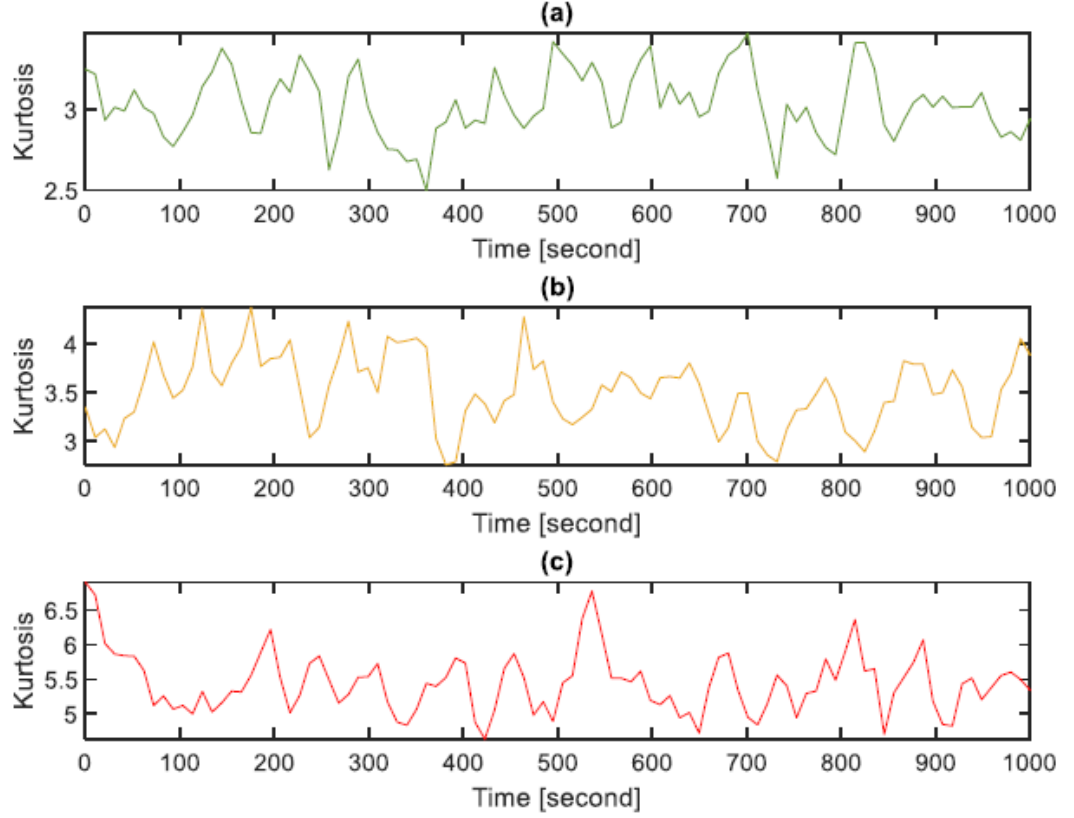


**Fig. 3.** Classification of the vibration signal into three-mode according to the presence or absence of faults. (a) Normal Signal; (b) Degraded Signal; (c) Faulty Signal.

**Kurtosis:** Kurtosis is increasingly used for fault detection of electrical machines, due to the simplicity of the algorithm, and its ability to detect non-stationary events. Kurtosis indicates the main peaks of the amplitude distribution of the vibration signals (Eq. 3). Fig 4. shows the kurtosis of different types of signals. For example, a good bearing with no flaws that cause impulses to the

signal will have a kurtosis-value~3 and in general, a kurtosis-value above 4 is a sign of a bad condition

$$K(x) = \frac{\sum (x_t(i) - E[x_t(i)])^4}{(\sum (x_t(i) - E[x_t(i)])^2)^2} \quad (3)$$



**Fig. 4.** The kurtosis in the time domain of the signals. (a) Normal Signal; (b) Degraded Signal; (c) Faulty Signal.

### 3 Predictive maintenance of rotating machines

In this part of the article, we have set up a system, which explains how to extract characteristics from measured data to carry out monitoring and forecasts. Based on the extracted functionalities, dynamic models are generated, validated, and used to predict the time of breakdowns so that actions can be taken before the actual breakdowns occur.

### **3.1 Classical machine learning based approaches**

Supervised learning is a type of machine learning that uses a known dataset to make predictions. The training data set includes input data and response values. Supervised learning algorithms seek to create a model capable of predicting the response values of a new data set. Using larger training data sets and optimizing, the model's hyperparameters can often increase the predictive power of the model and allow it to generalize well for new data sets. A set of test data is often used to validate the model. The k-NN, Decision Trees, and Support Vector Machines (SVM) algorithms are widely exploited for Time Series Classification in conjunction with one or few similarity measures [13]. Below, we briefly explain these techniques.

### **3.2 Support Vector Machines (SVM)**

SVMs have supervised learning models that analyze data used for non-probabilistic classification or regression analysis. The SVM algorithm is a so-called linear classifier, which means that, in the perfect case, the data must be linearly separable. It allows you to find the best separator (line, plane, or hyperplane) that best separates the two classes. The SVM method is based on the theory of statistical learning introduced by Vapnik in the 1990s [14].

### **3.3 Decision Trees**

If providing comprehensible classification results is of interest, Decision Trees are usually recommended. This method uses a top-down approach in order to build a tree recursively. At each internal node, there is a split test evaluation, and each leaf node contains information about a class to be assigned to the new test instance [15].

### **3.4 k-Nearest Neighbour (k-NN) Classifier**

The k-NN is one of the most commonly used algorithms in the Time Series Classification task is the k-NN classifier [16]. It is a simple, robust, and accurate classifier, which depends on very few parameters and requires no training. For distance measures, there are several distance functions; and all of them are used with continuous variables only. The k-NN classifier is simple but very effective with high accuracy.

### **3.5 Naive Bayes classifier**

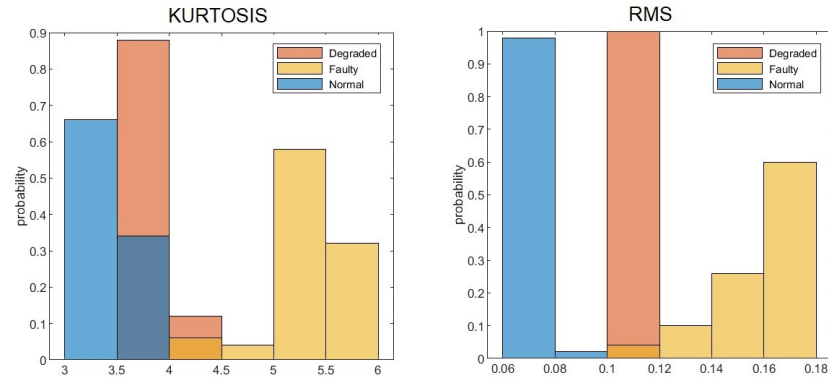
Naive Bayes classifier is utilizing Bayesian methods to classify data. During the training phase, the Naive Bayes classifier calculates the probability of every class based on independent variables. After that, these probabilities will be used to classify unlabeled data to the most likely class [17].

## **4 Results and Discussion**

With MATLAB's Diagnostic Feature Designer toolbox, we imported and visualized all of the vibration bearing signals in order to extract the essential characteristics of each data. Other signals

were generated randomly to emulate a faulty bearing vibration signal. Machine learning algorithms were used to classify those signals into different classes.

Before looking at the details of the Time Series Classification algorithms, we discuss the problem of interpretable and comprehensible ML. Figure 5 shows the characteristics obtained with Kurtosis and RMS. We can observe that the RMS data are better because it distinguishes better the three groups of data according to the level of failure compared to the characteristics of Kurtosis.

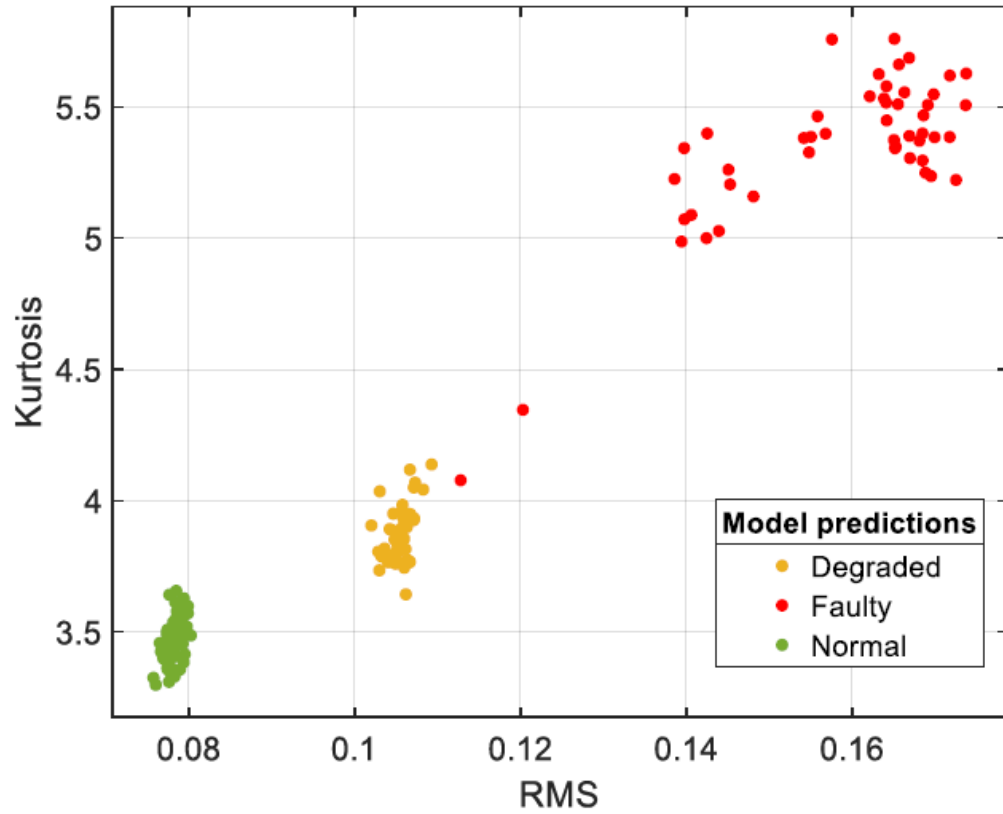


**Fig. 5.** Histogram of the characteristics of the vibration signals (Kurtosis. Vs. RMS).

#### 4.1 Test results with cross-validation

The Cross-Validation technique was used in this study to ensure the selection of the classifier models with the highest accuracy and the lowest prediction errors. As shown in Figure 6, the main dataset was divided into a training and validation dataset (85%) and a testing dataset (15%). The testing dataset was created randomly. The training and validation dataset was partitioned into  $k$  (a positive integer) equal size sub-dataset. Subsequently,  $k$  iterations of training and validation were performed. In each iteration cycle, a different sub-dataset was held-out for validation, while the remaining sub-datasets ( $k - 1$ ) were used for training. At the end of each training cycle, each classifier model of section 4 were run against the validation dataset to make predictions and then calculate the errors. Each model was stored in models array with its error rate. Upon completion, all observations were used for both training and validation; and the model with a lower error rate was selected as the final model.

In the Confusion Matrix analysis, the model accuracy is approximately 99.3% with the Naive Bayes, SVM, and k-NN methods, compared to the decision tree (98 %).



**Fig. 6.** Classification of signals with the cross-validation method.

## 5 Conclusion

This paper prospect of using vibrations to analyze the condition and possible diagnose mechanical presses was researched. The analysis was done by comparing some randomly chosen measurements from a different bearing signal and assigning them with one of three outputs, "normal," "degraded," or "faulty." The goal was to establish if the time domain classification RMS or statistical measure kurtosis would provide information that made each condition distinguishable. The following conclusions were made:

- Vibration analysis can be used in the predictive monitoring of rotating machines.
- RMS does seem to work for fault detection but should be complemented with some form of statistical kurtosis analysis.

The information obtained through the machine learning and monitoring system can help operators to prepare the necessary material before a failure occurs. Thus, traditional maintenance strategies involving corrective and preventive maintenance can be replaced by predictive maintenance (since predictive maintenance is the core of industry 4.0). With an excellent recognition rate of nearly 100%, this demonstrates that the use of recognition of vibration signals can allow a more precise prediction of the various elements that fail on machines.



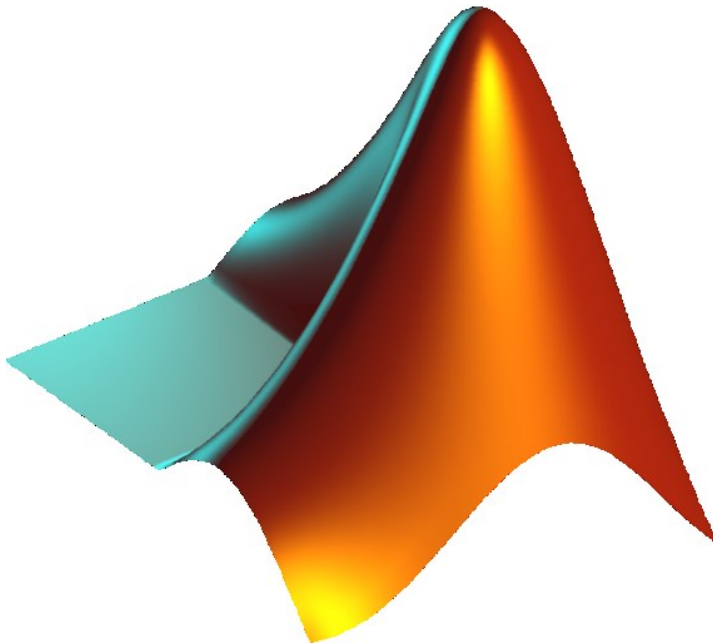
## References

1. Nembhard, A. D., Sinha, J. K., Elbhab, K., Pinkerton A. J.: Fault diagnosis of rotating machines using vibration and bearing temperature measurements. *Diagnostyka*, vol. 14 (3), 45-52 (2013).
2. Tabaszewski, M.: Optimization of a nearest neighbours classifier for diagnosis of condition of rolling bearings. *Diagnostyka*, vol. 15 (1), 37-42 (2014).
3. Astolfi, D., Castellani, F., Terzi, L.: Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm. *Diagnostyka*, vol. 15, 71-78 (2014).
4. Chehri, A., Jeon, G.: The Industrial Internet of Things: Examining How the IIoT Will Improve the Predictive Maintenance. *Lecture Notes of the Institute for Computer Sciences, Smart Innovation Systems and Technologies*, Springer, Heidelberg (2019).
5. Chehri, A., Jeon, G.: Routing Protocol in the Industrial Internet of Things for Smart Factory Monitoring. *Lecture Notes of the Institute for Computer Sciences, Smart Innovation Systems and Technologies*, Springer (2019).
6. Jeon, G., Awais, A., Chehri, A., Cuomo, S.: Special issue on video and imaging systems for critical engineering applications. *Multimedia Tools and Applications*, Springer (2020).
7. Jeon, G., Chehri, A., Cuomo, S., Din, S., Jabbar, S.: Special Issue on Real-time Behavioral Monitoring in IoT Applications using Big Data Analytics. *Concurrency and Computation: Practice and Experience*, John Wiley and Sons (2019).
8. Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., Lim, M. H: Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review. *IEEE Access*, vol. 7, 122644-122662 (2019).
9. Zhang, S., Zhang, S., Wang, B., Habetler, T. G.: Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review (2019).
10. Jaafar, A.: *Vibration Analysis and Diagnostics Guide*. College of Engineering, University of Basrah (2012).
11. Yang, W., Court, R., Jiang, S.: Wind turbine condition monitoring by the approach of SCADA data analysis. *Renew. Energy* 53 (2013).
12. Lee, J., Qiu, H., Yu, G., Lin, L.: *Bearing Data Set*. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, University of Cincinnati (2004).
13. Saber, M., Saadane, R., Aroussi, H., Chehri, A.: An Optimized Spectrum Sensing Implementation based on SVM, KNN and TREE Algorithms. *IEEE 15th International Conference on Signal Image Technology & Internet Based Systems*, Sorrento (NA), Italy (2019).
14. Boser, B., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM, Pittsburgh, Pennsylvania, United States, 144–152 (1992).
15. Yamada, Y., Suzuki, E., Yokoi, H., Takabayashi, K.: Decision-tree induction from timeseries data based on a standard-example splittest. *Machine Learning, Proceedings of the Twentieth International Conference*, Washington, DC, USA (2003).
16. Kilian, Q., Weinberger, J., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems*, 1473–1480 (2006).
17. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. *Machine Learning*. vol. 29, 131-163 (1997).

## ANNEXE 2 : PLATEFORME DE DÉVELOPPEMENT MATLAB

MathWorks a été fondée en 1984 par Jack Little et Cleve Moler, qui ont reconnu le besoin, parmi les ingénieurs et les scientifiques, d'environnements de calcul plus puissants et plus productifs que ceux fournis par des langages tels que Fortran et C. Ils ont combiné leur expertise en mathématiques, en ingénierie et en informatique pour développer MATLAB.

MATLAB ® est l'environnement informatique le plus simple et le plus productif pour les ingénieurs et les scientifiques. Avec les mathématiques, les graphiques et la programmation, il est conçu pour votre façon de penser et votre travail. (Source MathWorks)



### ANNEXE 3 : CALCULS STATISTIQUES

- Exemples d'échantillon de donnée utilisée pour la formation des algorithmes.

RMS	KURTOSIS	ÉTAT DU SIGNAL
0.0771829730453877	3.33603483085684	'Signal nominal'
0.0776099066319625	3.58306245144547	'Signal nominal'
0.0753548386567473	3.38734332480521	'Signal nominal'
0.0765868073658906	3.32951908124493	'Signal nominal'
0.0740066815633268	3.34657618325510	'Signal nominal'
0.0772364486116537	3.41210639684664	'Signal nominal'
0.0763822284580571	3.22272780378916	'Signal nominal'
0.0765719723094515	3.36811074688726	'Signal nominal'
0.0762263158017175	3.31402076871705	'Signal nominal'
0.0761112842349325	3.23864300736855	'Signal nominal'
0.0762350372742295	3.48025107600342	'Signal nominal'
0.0770855250603389	3.40568717581886	'Signal nominal'
0.0774851308820767	3.66705764590003	'Signal nominal'
0.0761065181151353	3.30226053567926	'Signal nominal'
0.0763500355578668	3.44313428353986	'Signal nominal'
0.0770615609735979	3.36952452332889	'Signal nominal'
0.0769959576226856	3.39192779691948	'Signal nominal'
0.0772802709829821	3.48063870982931	'Signal nominal'
0.0778043576312167	3.37357556775777	'Signal nominal'
0.104292161878529	3.78059286308214	'Signal dégradé'
0.105857970701020	3.93244374633339	'Signal dégradé'
0.106209899142131	3.76181316705386	'Signal dégradé'
0.103015902045360	4.03529558977095	'Signal dégradé'
0.105969381264709	3.74462243622083	'Signal dégradé'
0.105959381036735	3.85442947960545	'Signal dégradé'
0.104640304193552	3.95032983484248	'Signal dégradé'
0.105642071937660	3.89518124644077	'Signal dégradé'
0.104359543672197	3.78163487880054	'Signal dégradé'
0.106627871719097	3.76786425607960	'Signal dégradé'
0.106621774203844	3.76546977659139	'Signal dégradé'
0.104702685465801	3.78698191517117	'Signal dégradé'
0.104210103183276	3.89163867404976	'Signal dégradé'
0.106123280361879	3.81499632203021	'Signal dégradé'
0.106856370132487	3.92595116931718	'Signal dégradé'
0.104840582571381	3.77730123365145	'Signal dégradé'
0.103599359765157	3.81763520109115	'Signal dégradé'
0.106160850007049	3.64305166141831	'Signal dégradé'
0.106318528451949	3.77469463228168	'Signal dégradé'
0.104912635482981	3.75948687212536	'Signal dégradé'
0.169938527344955	5.25506000847602	'Signal défaillant'

0.175521273583772	5.06302987928288	'Signal défaillant'
0.169596059904576	5.11457028262786	'Signal défaillant'
0.184598116457374	5.41082616164473	'Signal défaillant'
0.179653305442039	5.00814604845825	'Signal défaillant'
0.171568919221895	4.31351159938339	'Signal défaillant'
0.169244800482126	4.07642680595926	'Signal défaillant'
0.157969592294985	3.71120633020862	'Signal défaillant'
0.150929011148453	3.87865141483164	'Signal défaillant'
0.158335274128044	3.50930371132582	'Signal défaillant'
0.170505698333114	3.35985567966641	'Signal défaillant'
0.167702628137949	3.39780194254201	'Signal défaillant'
0.179620514824180	3.53024267579308	'Signal défaillant'
0.192045559881799	3.75710494704804	'Signal défaillant'
0.199996212854772	3.86093588298128	'Signal défaillant'
0.224457285246864	4.43136806528024	'Signal défaillant'
0.265929517726854	5.18835589883582	'Signal défaillant'
0.201158112757419	4.81622480425710	'Signal défaillant'
0.196909587859547	4.42840919081935	'Signal défaillant'
0.195597599044452	4.19071256455968	'Signal défaillant'
0.166472801539303	3.70693390108899	'Signal défaillant'
0.162400464708387	3.83651464086740	'Signal défaillant'
0.170069323239177	3.80797636666098	'Signal défaillant'
0.163082954650867	3.23859552164685	'Signal défaillant'
0.197988164621219	3.69187524643936	'Signal défaillant'
0.188476326845710	3.64315764452921	'Signal défaillant'
0.215437279359660	4.45871159594587	'Signal défaillant'
0.183851038866220	3.88557766692968	'Signal défaillant'
0.163875821770442	3.26447976532175	'Signal défaillant'
0.158321338241495	3.32690805757469	'Signal défaillant'
0.161203488931691	3.59627831172139	'Signal défaillant'

- Tableau de classement des caractéristiques.

